Chapter 14

# Spatial survival models

The use of survival models involving a random effect or "frailty" term is becoming more common. Usually the random effects are assumed to represent different clusters, and clusters are assumed to be independent. In this chapter, we consider random effects corresponding to clusters that are spatially arranged, such as clinical sites or geographical regions. That is, we might suspect that random effects corresponding to strata in closer proximity to each other might also be similar in magnitude.

Survival models have a long history in the biostatistical and medical literature (see, e.g., Cox and Oakes, 1984). Very often, time-to-event data will be grouped into *strata* (or *clusters*), such as clinical sites, geographic regions, and so on. In this setting, a hierarchical modeling approach using stratum-specific parameters called *frailties* is often appropriate. Introduced by Vaupel, Manton, and Stallard (1979), this is a mixed model with random effects (the frailties) that correspond to a stratum's overall health status.

To illustrate, let $t_{ij}$ be the time to death or censoring for subject $j$ in stratum $i$, $j = 1, \ldots, n_i$, $i = 1, \ldots, I$. Let $\mathbf{x}_{ij}$ be a vector of individual-specific covariates. The usual assumption of proportional hazards $h(t_{ij}; \mathbf{x}_{ij})$ enables models of the form

$$h(t_{ij}; \mathbf{x}_{ij}) = h_0(t_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \,, \tag{14.1}$$

where $h_0$ is the *baseline hazard*, which is affected only multiplicatively by the exponential term involving the covariates. In the frailty setting, model (14.1) is extended to

$$\begin{aligned} h(t_{ij}; x_{ij}) &= h_0(t_{ij})\, \omega_i \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij}) \\ &= h_0(t_{ij}) \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i) \,, \end{aligned} \tag{14.2}$$

where $W_i \equiv \log \omega_i$ is the stratum-specific frailty term, designed to capture differences among the strata. Typically a simple i.i.d. specification for the $W_i$ is assumed, e.g.,

$$W_i \stackrel{iid}{\sim} N(0, \sigma^2) \,. \tag{14.3}$$

With the advent of MCMC computational methods, the Bayesian approach to fitting hierarchical frailty models such as these has become increasingly popular (see, e.g., Carlin and Louis, 2000, Sec. 7.6). Perhaps the simplest approach is to assume a *parametric* form for the baseline hazard $h_0$. While a variety of choices (gamma, lognormal, etc.) have been explored in the literature, in Section 14.1 we adopt the Weibull, which seems to represent a good tradeoff between simplicity and flexibility. This then produces

$$h(t_{ij}; x_{ij}) = \rho t_{ij}^{\rho - 1} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i) \,. \tag{14.4}$$

Now, placing prior distributions on $\rho, \boldsymbol{\beta}$, and $\sigma^2$ completes the Bayesian model specification. Such models are by now a standard part of the literature, and easily fit (at least in the univariate case) using `WinBUGS`. Carlin and Hodges (1999) consider further extending model

(14.4) to allow stratum-specific baseline hazards, i.e., by replacing $\rho$ by $\rho_i$. MCMC fitting is again routine given a distribution for these new random effects, say, $\rho_i \overset{iid}{\sim} Gamma(\alpha, 1/\alpha)$, so that the $\rho_i$ have mean 1 (corresponding to a constant hazard over time) but variance $1/\alpha$.

A richer but somewhat more complex alternative is to model the baseline hazard *nonparametrically*. In this case, letting $\gamma_{ij}$ be a death indicator (0 if alive, 1 if dead) for patient $ij$, we may write the likelihood for our model $L(\beta, \mathbf{W}; \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma})$ generically as

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left\{h\left(t_{ij};\mathbf{x}_{ij}\right)\right\}^{\gamma_{ij}}\exp\left\{-H_{0i}\left(t_{ij}\right)\exp\left(\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i\right)\right\} ,$$

where $H_{0i}(t) = \int_0^t h_{0i}(u)\,du$, the integrated baseline hazard. A frailty distribution parametrized by $\lambda$, $p(\mathbf{W}|\lambda)$, coupled with prior distributions for $\lambda$, $\boldsymbol{\beta}$, and the hazard function $h$ complete the hierarchical Bayesian model specification.

In this chapter we consider both parametric and semiparametric hierarchical survival models for data sets that are spatially arranged. Such models might be appropriate anytime we suspect that frailties $W_i$ corresponding to strata in closer proximity to each other might also be similar in magnitude. This could arise if, say, the strata corresponded to hospitals in a given region, to counties in a given state, and so on. The basic assumption here is that "expected" survival times (or hazard rates) will be more similar in proximate regions, due to underlying factors (access to care, willingness of the population to seek care, etc.) that vary spatially. We hasten to remind the reader that this does not imply that the observed survival times from subjects in proximate regions must be similar, since they include an extra level of randomness arising from their variability around their (spatially correlated) underlying model quantities.

## 14.1 Parametric models

### *14.1.1 Univariate spatial frailty modeling*

While it is possible to identify centroids of geographic regions and employ spatial process modeling for these locations, the effects in our examples are more naturally associated with areal units. As such we work exclusively with CAR models for these effects, i.e., we assume that

$$\mathbf{W} \mid \lambda \sim CAR(\lambda) . \tag{14.5}$$

Also, we note that the resulting model for, say, (14.2) is an extended example of a generalized linear model for areal spatial data (Section 6.5). That is, (14.2) implies that

$$f(t_{ij}|\boldsymbol{\beta}, x_{ij}, W_i) = h_0(t_{ij})e^{\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i}e^{-H_0(t_{ij})\exp(\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i)} . \tag{14.6}$$

In other words, $U_{ij} = H_0(t_{ij}) \sim \text{Exponential}\Big(\exp[-(\boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i)]\Big)$ so $-\log EH_0(t_{ij}) = \boldsymbol{\beta}^T\mathbf{x}_{ij}+W_i$. The analogy with (6.29) and $g(\eta_i)$ is clear. The critical difference is that in Section 6.5 the link $g$ is assumed known; here the link to the linear scale requires $h_0$, which is unknown (and will be modeled parametrically or nonparametrically).

Finally, we remark that it would certainly be possible to include both spatial and nonspatial frailties, which as already seen (Subsection 6.4.3) is now common practice in areal data modeling. Here, this would mean supplementing our spatial frailties $W_i$ with a collection of nonspatial frailties, say, $V_i \overset{iid}{\sim} N(0, 1/\tau)$. The main problem with this approach is again that the frailties now become identified only by the prior, and so the proper choice of priors for $\tau$ and $\lambda$ (or $\boldsymbol{\theta}$) becomes problematic. Another problem is the resultant decrease in algorithm performance wrought by the addition of so many additional, weakly identified parameters.

*14.1.1.1 Bayesian implementation*

As already mentioned, the models outlined above are straightforwardly implemented in a Bayesian framework using MCMC methods. In the parametric case, say (14.4), the joint posterior distribution of interest is

$$p\left(\boldsymbol{\beta}, \mathbf{W}, \rho, \lambda \mid \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma}\right) \propto L\left(\boldsymbol{\beta}, \mathbf{W}, \rho\,;\, \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma}\right) p\left(\mathbf{W}|\lambda\right) p(\boldsymbol{\beta}) p\left(\rho\right) p(\lambda)\,, \tag{14.7}$$

where the first term on the right-hand side is the Weibull likelihood, the second is the CAR distribution of the random frailties, and the remaining terms are prior distributions. In (14.7), $\mathbf{t} = \{t_{ij}\}$ denotes the collection of times to death, $\mathbf{x} = \{\mathbf{x}_{ij}\}$ the collection of covariate vectors, and $\boldsymbol{\gamma} = \{\gamma_{ij}\}$ the collection of death indicators for all subjects in all strata.

For our investigations, we retain the parametric form of the baseline hazard given in (14.4). Thus $L\left(\boldsymbol{\beta}, \mathbf{W}, \rho\,;\, \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma}\right)$ is proportional to

$$\prod_{i=1}^{I} \prod_{j=1}^{n_i} \left\{\rho t_{ij}^{\rho-1} \exp\left(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i\right)\right\}^{\gamma_{ij}} \exp\left\{-t_{ij}^{\rho} \exp\left(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i\right)\right\}. \tag{14.8}$$

The model specification in the Bayesian setup is completed by assigning prior distributions for $\boldsymbol{\beta}, \rho$, and $\lambda$. Typically, a flat (improper uniform) prior is chosen for $\boldsymbol{\beta}$, while vague but proper priors are chosen for $\rho$ and $\lambda$, such as a $G(\alpha, 1/\alpha)$ prior for $\rho$ and a $G(a, b)$ prior for $\lambda$. Hence the only extension beyond the disease mapping illustrations of Section 6.4 is the need to update $\rho$.

**Example 14.1** *(Application to Minnesota infant mortality data).* We apply the methodology above to the analysis of infant mortality in Minnesota, originally considered by Banerjee, Wall, and Carlin (2003). The data were obtained from the linked birth-death records data registry kept by the Minnesota Department of Health. The data comprise 267,646 live births occurring during the years 1992–1996 followed through the first year of life, together with relevant covariate information such as birth weight, sex, race, mother's age, and the mother's total number of previous births. Because of the careful linkage connecting infant death certificates with birth certificates (even when the death occurs in a separate state), we assume that each baby in the data set that is not linked with a death must have been alive at the end of one year. Of the live births, only 1,547 babies died before the end of their first year. The number of days they lived is treated as the response $t_{ij}$ in our models, while the remaining survivors were treated as "censored," or in other words, alive at the end of the study period. In addition to this information, the mother's Minnesota county of residence prior to the birth is provided. We implement the areal frailty model (14.5), the nonspatial frailty model (14.3), and a simple nonhierarchical ("no-frailty") model that sets $W_i = 0$ for all $i$.

For all of our models, we adopt a flat prior for $\boldsymbol{\beta}$, and a $G(\alpha, 1/\alpha)$ prior for $\rho$, setting $\alpha = 0.01$. Metropolis random walk steps with Gaussian proposals were used for sampling from the full conditionals for $\boldsymbol{\beta}$, while Hastings independence steps with gamma proposals were used for updating $\rho$. As for $\lambda$, in our case we are fortunate to have a data set that is large relative to the number of random effects to be estimated. As such, we simply select a vague (mean 1, variance 1000) gamma specification for $\lambda$, and rely on the data to overwhelm the priors.

Table 14.1 compares our three models in terms of two of the criteria discussed in Subsection 5.2.3, DIC and effective model size $p_D$. For the no-frailty model, we see a $p_D$ of 8.72, very close to the actual number of parameters, 9 (8 components of $\boldsymbol{\beta}$ plus the Weibull parameter $\rho$). The random effects models have substantially larger $p_D$ values, though much smaller than their actual parameter counts (which would include the 87 random frailties

| Model | $p_D$ | DIC |
|---|---|---|
| No-frailty | 8.72 | 511 |
| Nonspatial frailty | 39.35 | 392 |
| CAR frailty | 34.52 | 371 |

Table 14.1 *DIC and effective number of parameters $p_D$ for competing parametric survival models.*

| Covariate | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Intercept | −2.135 | −2.024 | −1.976 |
| Sex (boys = 0) | | | |
| girls | −0.271 | −0.189 | −0.105 |
| Race (white = 0) | | | |
| black | −0.209 | −0.104 | −0.003 |
| Native American | 0.457 | 0.776 | 1.004 |
| unknown | 0.303 | 0.871 | 1.381 |
| Mother's age | −0.005 | −0.003 | −0.001 |
| Birth weight in kg | −1.820 | −1.731 | −1.640 |
| Total births | 0.064 | 0.121 | 0.184 |
| $\rho$ | 0.411 | 0.431 | 0.480 |
| $\sigma$ | 0.083 | 0.175 | 0.298 |

Table 14.2 *Posterior summaries for the nonspatial frailty model.*

$W_i$); apparently there is substantial shrinkage of the frailties toward their grand mean. The DIC values suggest that each of these models is substantially better than the no-frailty model, despite their increased size. Though the spatial frailty model has the best DIC value, plots of the full estimated posterior deviance distributions (not shown) suggest substantial overlap. On the whole we seem to have modest support for the spatial frailty model over the ordinary frailty model.

Tables 14.2 and 14.3 provide 2.5, 50, and 97.5 posterior percentiles for the main effects in our two frailty models. In both cases, all of the predictors are significant at the .05 level. Since the reference group for the sex variable is boys, we see that girls have a lower hazard of death during the first year of life. The reference group for the race variables is white; the Native American beta coefficient is rather striking. In the CAR model, this covariate increases the posterior median hazard rate by a factor of $e^{0.782} = 2.19$. The effect of "unknown" race is also significant, but more difficult to interpret: in this group, the race of the infant was not recorded on the birth certificate. Separate terms for Hispanics, Asians, and Pacific Islanders were also originally included in the model, but were eliminated after emerging as not significantly different from zero. Note that the estimate of $\rho$ is quite similar across models, and suggests a decreasing baseline hazard over time. This is consistent with the fact that a high proportion (495, or 32%) of the infant deaths in our data set occurred in the first *day* of life: the force of mortality (hazard rate) is very high initially, but drops quickly and continues to decrease throughout the first year.

A benefit of fitting the spatial CAR structure is seen in the reduction of the length of the 95% credible intervals for the covariates in the spatial models compared to the i.i.d. model. As we might expect, there are modest efficiency gains when the model that better specifies the covariance structure of its random effects is used. That is, since the spatial dependence priors for the frailties are in better agreement with the likelihood than is the independence prior, the prior-to-posterior learning afforded by Bayes' Rule leads to smaller posterior variances in the former cases. Most notably, the 95% credible set for the effect of "unknown" race is (0.303, 1.381) under the nonspatial frailty model (Table 14.2), but

| Covariate | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Intercept | −2.585 | −2.461 | −2.405 |
| Sex (boys = 0) | | | |
| girls | −0.224 | −0.183 | −0.096 |
| Race (white = 0) | | | |
| black | −0.219 | −0.105 | −0.007 |
| Native American | 0.455 | 0.782 | 0.975 |
| unknown | 0.351 | 0.831 | 1.165 |
| Mother's age | −0.005 | −0.004 | −0.003 |
| Birth weight in kg | −1.953 | −1.932 | −1.898 |
| Total births | 0.088 | 0.119 | 0.151 |
| $\rho$ | 0.470 | 0.484 | 0.497 |
| $\lambda$ | 12.62 | 46.07 | 100.4 |

Table 14.3 *Posterior summaries for the CAR frailty model.*

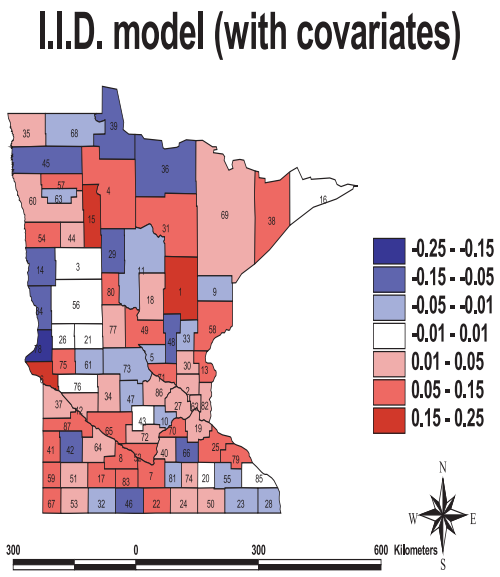


Figure 14.1 *Posterior median frailties, i.i.d. model with covariates, Minnesota county-level infant mortality data.*

(0.351, 1.165) under the CAR frailty model (Table 14.3), a reduction in length of roughly 25%.

Figures 14.1 and 14.2 map the posterior medians of the $W_i$ under the nonspatial (i.i.d. frailties) and CAR models, respectively, where the models include all of the covariates listed in Tables 14.2 and 14.3. As expected, no clear spatial pattern is evident in the i.i.d. map, but from the CAR map we are able to identify two clusters of counties having somewhat higher hazards (in the southwest following the Minnesota River, and in the northeast "arrowhead" region), and two clusters with somewhat lower hazards (in the northwest, and the southeastern corner). Thus, despite the significance of the covariates now in these models, Figure 14.2 suggests the presence of some still-missing, spatially varying covariate(s) relevant for infant mortality. Such covariates might include location of birth (home or hospital),

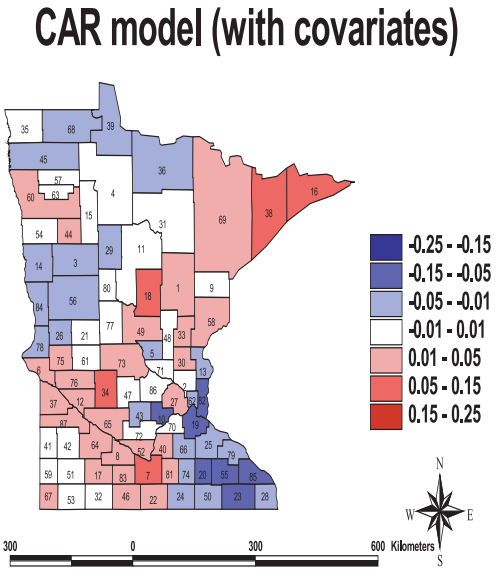


Figure 14.2 *Posterior median frailties, CAR model with covariates, Minnesota county-level infant mortality data.*

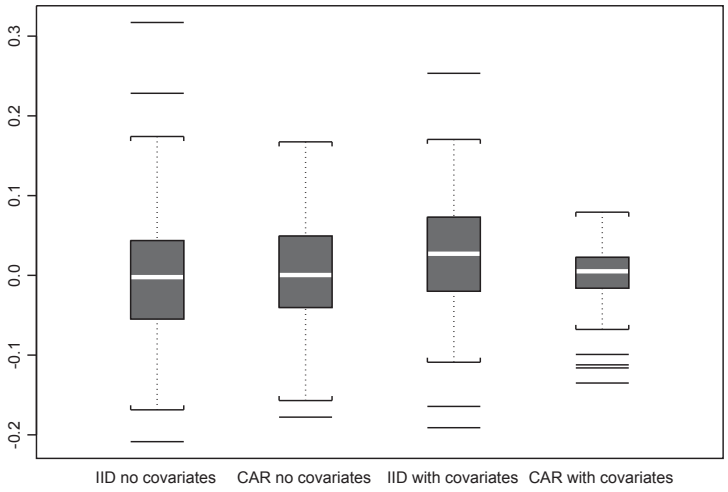


Figure 14.3 *Boxplots of posterior median frailties, i.i.d. and CAR models with and without covariates.*

overall quality of available health or hospital care, mother's economic status, and mother's number of prior abortions or miscarriages.

In addition to the improved appearance and epidemiological interpretation of Figure 14.2, another reason to prefer the CAR model is provided in Figure 14.3, which shows boxplots of the posterior median frailties for the two cases corresponding to Figures 14.1 and 14.2, plus two preliminary models in which *no* covariates $\mathbf{x}$ are included. The tightness of the full CAR boxplot suggests this model is best at reducing the need for the frailty terms. This is as it should be, since these terms are essentially spatial residuals, and rep-

resent lingering lack of fit in our spatial model (although they may well also account for some excess *non*spatial variability, since our current models do not include nonspatial frailty terms). Note that all of the full CAR residuals are in the range (–0.15, 0.10), or (0.86, 1.11) on the hazard scale, suggesting that missing spatially varying covariates have only a modest (10 to 15%) impact on the hazard; from a practical standpoint, this model fits quite well. ■

### *14.1.2 Spatial frailty versus logistic regression models*

In many contexts (say, a clinical trial enrolling and following patients at spatially proximate clinical centers), a spatial survival model like ours may be the only appropriate model. However, since the Minnesota infant mortality data does not have any babies censored because of loss to followup, competing risks, or any reason other than the end of the study, there is no ambiguity in defining a *binary* survival outcome for use in a random effects logistic regression model. That is, we replace the event time data $t_{ij}$ with an indicator of whether the subject did ($Y_{ij} = 0$) or did not ($Y_{ij} = 1$) survive the first year. Letting $p_{ij} = Pr(Y_{ij} = 1)$, our model is then

$$logit(p_{ij}) = \widetilde{\boldsymbol{\beta}}^T \mathbf{x}_{ij} + \widetilde{W}_i \,, \tag{14.9}$$

with the usual flat prior for $\widetilde{\boldsymbol{\beta}}$ and an i.i.d. or CAR prior for the $\widetilde{W}_i$. As a result, (14.9) is exactly an example of a generalized linear model for areal spatial data.

Other authors (Doksum and Gasko, 1990; Ingram and Kleinman, 1989) have shown that in this case of no censoring before followup (and even in cases of equal censoring across groups), it is possible to get results for the $\widetilde{\boldsymbol{\beta}}$ parameters in the logistic regression model very similar to those obtained in the proportional hazards model (14.1), except of course for the differing interpretations (log odds versus log relative risk, respectively). Moreover when the probability of death is very small, as it is in the case of infant mortality, the log odds and log relative risk become even more similar. Since it uses more information (i.e., time to death rather than just a survival indicator), intuitively, the proportional hazards model should make gains over the logistic model in terms of power to detect significant covariate effects. Yet, consistent with the simulation studies performed by Ingram and Kleinman (1989), our experience with the infant mortality data indicate that only a marginal increase in efficiency (decrease in variance) is exhibited by the posterior distributions of the parameters.

On the other hand, we did find some difference in terms of the estimated random effects in the logistic model compared to the proportional hazards model. Figure 14.4 shows a scatterplot of the estimated posterior medians of $W_i$ versus $\widetilde{W}_i$ for each county obtained from the models where there were no covariates, and the random effects were assumed to i.i.d. The sample correlation of these estimated random effects is 0.81, clearly indicating that they are quite similar. Yet there are still some particular counties that result in rather different values under the two models. One way to explain this difference is that the hazard functions are not exactly proportional across the 87 counties of Minnesota. A close examination of the counties that had differing $\widetilde{W}_i$ versus $W_i$ shows that they had different average times at death compared to other counties with similar overall death rates. Consider for example County 70, an outlier circled in Figure 14.4, and its comparison to circled Counties 73, 55, and 2, which have similar death rates (and hence roughly the same horizontal position in Figure 14.4). We find County 70 has the smallest mean age at death, implying that it has more early deaths, explaining its smaller frailty estimate. Conversely, County 14 has a higher average time at death but overall death rates similar to Counties 82, 48, and 5 (again note the horizontal alignment in Figure 14.4), and as a result has higher estimated frailty. A lack of proportionality in the baseline hazard rates across counties thus appears to manifest as a departure from linearity in Figure 14.4.

We conclude this subsection by noting that previous work by Carlin and Hodges (1999)

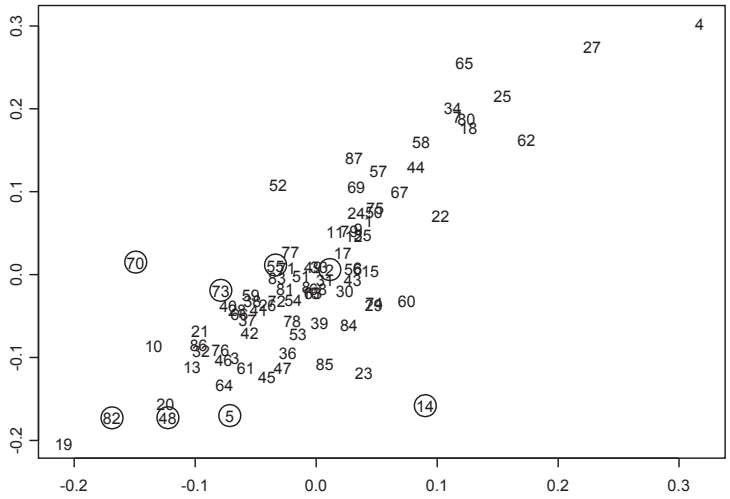


Figure 14.4 *Posterior medians of the frailties $W_i$ (horizontal axis) versus posterior medians of the logistic random effects $\widetilde{W}_i$ (vertical axis). Plotting character is county number; significance of circled counties is described in the text.*

suggests a generalization of our basic model (14.4) to

$$h(t_{ij}; x_{ij}) = \rho_i t_{ij}^{\rho_i - 1} \exp(\boldsymbol{\beta}^T \mathbf{x}_{ij} + W_i) \; .$$

That is, we allow two sets of random effects: the existing frailty parameters $W_i$, and a new set of shape parameters $\rho_i$. This then allows both the overall level and the shape of the hazard function over time to vary from county to county. Either i.i.d. or CAR priors could be assigned to these two sets of random effects, which could themselves be correlated within county. In the latter case, this might be fit using the MCAR model of Section 10.1; see Jin and Carlin (2003), as well as Section 14.4.

### 14.2 Semiparametric models

While parametric models are easily interpretable and often afford a surprisingly good fit to survival data, many practitioners continue to prefer the additional richness of the nonparametric baseline hazard offered by the celebrated Cox model. In this section we turn to nonparametric models for the baseline hazard. Such models are often referred to as *semiparametric*, since we continue to assume proportional hazards of the form (14.1) in which the covariate effects are still modeled parametrically, While Li and Ryan (2002) address this problem from a classical perspective, in this section we follow the hierarchical Bayesian approach of Banerjee and Carlin (2002).

Within the Bayesian framework, several authors have proposed treating the Cox partial likelihood as a full likelihood, to obtain a posterior distribution for the treatment effect. However, this approach does not allow fully hierarchical modeling of stratum-specific baseline hazards (with stratum-specific frailties) because the baseline hazard is implicit in the partial likelihood computation. In the remainder of this section, we describe two possible methodological approaches to modeling the baseline hazard in Cox regression, which thus lead to two semiparametric spatial frailty techniques. We subsequently revisit the Minnesota infant mortality data.

*14.2.1 Beta mixture approach*

Our first approach uses an idea of Gelfand and Mallick (1995) that flexibly models the integrated baseline hazard as a mixture of monotone functions. In particular, these authors use a simple transformation to map the integrated baseline hazard onto the interval $[0, 1]$, and subsequently approximate this function by a weighted mixture of incomplete beta functions. Implementation issues are discussed in detail by Gelfand and Mallick (1995) and also by Carlin and Hodges (1999) for stratum-specific baseline hazards. The likelihood and Bayesian hierarchical setup remain exactly as above.

Thus, we let $h_{0i}(t)$ be the baseline hazard in the $i$th region and $H_{0i}(t)$ be the corresponding integrated baseline hazard, and define

$$J_{0i}(t) = a_0 H_{0i}(t) / [a_0 H_{0i}(t) + b_0] \ ,$$

which conveniently takes values in $[0, 1]$. We discuss below the choice of $a_0$ and $b_0$ but note that this is not as much a modeling issue as a computational one, important only to ensure appropriate coverage of the interval $[0, 1]$. We next model $J_{0i}(t)$ as a mixture of $Beta(r_l, s_l)$ cdfs, for $l = 1, \ldots, m$. The $r_l$ and $s_l$ are chosen so that the beta cdfs have evenly spaced means and are centered around $\widetilde{J_0}(t)$, a suitable function transforming the time scale to $[0, 1]$. We thus have

$$J_{0i}(t) = \sum_{l=1}^{m} v_{il}\, IB\left(\widetilde{J_0}(t); r_l, s_l\right) \ ,$$

where $\sum_{l=1}^{m} v_{il} = 1$ for all $i$, and $IB(\cdot; a, b)$ denotes the incomplete beta function (i.e., the cdf of a $Beta(a, b)$ distribution). Since any distribution function on $[0, 1]$ can be approximated arbitrarily well by a finite mixture of beta cdfs, the same is true for $J_{0i}$, an increasing function that maps $[0, 1]$ onto itself. Thus, working backward, we find the following expression for the cumulative hazard in terms of the above parameters:

$$H_{0i}(t) = \frac{b_0 \sum_{l=1}^{m} v_{il}\, IB\left(\widetilde{J_0}(t); r_l, s_l\right)}{a_0 \left\{1 - \sum_{l=1}^{m} v_{il}\, IB\left(\widetilde{J_0}(t); r_l, s_l\right)\right\}} \ .$$

Taking derivatives, we have for the hazard function,

$$h_{0i}(t) = \frac{b_0 \frac{\partial}{\partial t}\widetilde{J_0}(t) \sum_{l=1}^{m} v_{il} Beta\left(\widetilde{J_0}(t); r_l, s_l\right)}{a_0 \left\{1 - \sum_{l=1}^{m} v_{il}\, IB\left(\widetilde{J_0}(t); r_l, s_l\right)\right\}^2} \ .$$

Typically $m$, the number of mixands of the beta cdfs, is fixed, as are the $\{(r_l, s_l)\}_{l=1}^{m}$, so chosen that the resulting beta densities cover the interval $[0, 1]$. For example, we might fix $m = 5$, $\{r_l\} = (1, 2, 3, 4, 5)$ and $\{s_l\} = (5, 4, 3, 2, 1)$, producing five evenly-spaced beta cdfs.

Regarding the choice of $a_0$ and $b_0$, we note that it is intuitive to specify $\widetilde{J_0}(t)$ to represent a plausible central function around which the $J_{0i}$'s are distributed. Thus, if we consider the cumulative hazard function of an exponential distribution to specify $\widetilde{J_0}(t)$, then we get $\widetilde{J_0}(t) = a_0 t/(a_0 t + b_0)$. In our Minnesota infant mortality data set, since the survival times ranged between 1 day and 365 days, we found $a_0 = 5$ and $b_0 = 100$ lead to values for $\widetilde{J_0}(t)$ that largely cover the interval $[0, 1]$, and so fixed them as such. The likelihood is thus a function of the regression coefficients $\boldsymbol{\beta}$, the stratum-specific weight vectors $\mathbf{v}_i = (v_{i1}, ..., v_{im})^T$, and the spatial effects $W_i$. It is natural to model the $\mathbf{v}_i$'s as draws from a Dirichlet$(\phi_1, \ldots, \phi_m)$ distribution, where for simplicity we often take $\phi_1 = \cdots = \phi_m = \phi$.

*14.2.2 Counting process approach*

The second nonparametric baseline hazard modeling approach we investigate is that of Clayton (1991, 1994). While the method is less transparent theoretically, it is gaining popularity among Bayesian practitioners due to its ready availability within `WinBUGS`. Here we give only the essential ideas, referring the reader to Andersen and Gill (1982) or Clayton (1991) for a more complete treatment. The underlying idea is that the number of failures up to time $t$ is assumed to arise from a *counting process* $N(t)$. The corresponding *intensity process* is defined as

$$I\left(t\right)dt = E\left(dN\left(t\right)|F_{t-}\right) \ ,$$

where $dN(t)$ is the increment of $N$ over the time interval $[t, t+dt)$, and $F_{t-}$ represents the available data up to time $t$. For each individual, $dN\left(t\right)$ therefore takes the value 1 if the subject fails in that interval, and 0 otherwise. Thus $dN\left(t\right)$ may be thought of as the "death indicator process," analogous to $\gamma$ in the model of the previous subsection. For the $j$th subject in the $i$th region, under the proportional hazards assumption, the intensity process (analogous to our hazard function $h\left(t_{ij};\mathbf{x}_{ij}\right)$) is modeled as

$$I_{ij}(t) = Y_{ij}(t)\lambda_0\left(t\right)\exp\left(\boldsymbol{\beta}^T\mathbf{x}_{ij} + W_i\right) \ ,$$

where $\lambda_0\left(t\right)$ is the baseline hazard function and $Y_{ij}\left(t\right)$ is an indicator process taking the value 1 or 0 according to whether or not subject $i$ is observed at time $t$. Under the above formulation and keeping the same notation as above for $\mathbf{W}$ and $\mathbf{x}$, a Bayesian hierarchical model may be formulated as:

$$\begin{aligned} dN_{ij}\left(t\right) &\sim Poisson\left(I_{ij}\left(t\right)dt\right) \ , \\ I_{ij}\left(t\right)dt &= Y_{ij}(t)\exp\left(\boldsymbol{\beta}^T\mathbf{x}_{ij} + W_i\right)d\Lambda_0\left(t\right) \ , \\ d\Lambda_0\left(t\right) &\sim Gamma\left(c\,d\Lambda_0^*\left(t\right), c\right) \ . \end{aligned}$$

As before, priors $p\left(\mathbf{W}|\lambda\right)$, $p\left(\lambda\right)$, and $p(\boldsymbol{\beta})$ are required to completely specify the Bayesian hierarchical model. Here, $d\Lambda_0\left(t\right) = \lambda_0\left(t\right)dt$ may be looked upon as the increment or jump in the integrated baseline hazard function occurring during the time interval $[t, t+dt)$. Since the conjugate prior for the Poisson mean is the gamma distribution, $\Lambda_0\left(t\right)$ is conveniently modeled as a process whose increments $d\Lambda_0\left(t\right)$ are distributed according to gamma distributions. The parameter $c$ in the above setup represents the degree of confidence in our prior guess for $d\Lambda_0\left(t\right)$, given by $d\Lambda_0^*\left(t\right)$. Typically, the prior guess $d\Lambda_0^*\left(t\right)$ is modeled as $r\,dt$, where $r$ is a guess at the failure rate per unit time. The `LeukFr` example in the `WinBUGS` examples manual offers an illustration of how to code the above formulation.

**Example 14.2** *(Application to Minnesota infant mortality data, continued).* We now apply the methodology above to the reanalysis of our Minnesota infant mortality data set. For both the CAR and nonspatial models we implemented the Cox model with the two semiparametric approaches outlined above. We found very similar results, and so in our subsequent analysis we present only the results with the beta mixture approach (Subsection 14.2.1). For all of our models, we adopt vague Gaussian priors for $\boldsymbol{\beta}$. Since the full conditionals for each component of $\boldsymbol{\beta}$ are log-concave, adaptive rejection sampling was used for sampling from the $\boldsymbol{\beta}$ full conditionals. As in Section 14.1, we again simply select a vague $G(0.001, 1000)$ (mean 1, variance 1000) specification for CAR smoothness parameter $\lambda$, though we maintain more informative priors on the other variance components.

Table 14.4 compares our three models in terms of DIC and effective model size $p_D$. For the no-frailty model, we see a $p_D$ of 6.82, reasonably close to the actual number of parameters, 8 (the components of $\boldsymbol{\beta}$). The other two models have substantially larger $p_D$ values, though much smaller than their actual parameter counts (which would include the 87

| Model | $p_D$ | DIC |
|---|---|---|
| No-frailty | 6.82 | 507 |
| Nonspatial frailty | 27.46 | 391 |
| CAR frailty | 32.52 | 367 |

Table 14.4 *DIC and effective number of parameters $p_D$ for competing nonparametric survival models.*

| Covariate | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Intercept | −2.524 | −1.673 | −0.832 |
| Sex (boys = 0) | | | |
| girls | −0.274 | −0.189 | −0.104 |
| Race (white = 0) | | | |
| black | −0.365 | −0.186 | −0.012 |
| Native American | 0.427 | 0.737 | 1.034 |
| unknown | 0.295 | 0.841 | 1.381 |
| Mother's age | −0.054 | −0.035 | −0.014 |
| Birth weight in kg | −1.324 | −1.301 | −1.280 |
| Total births | 0.064 | 0.121 | 0.184 |

Table 14.5 *Posterior summaries for the nonspatial semiparametric frailty model.*

random frailties $W_i$); apparently there is substantial shrinkage of the frailties toward their grand mean. The DIC values suggest that both of these models are substantially better than the no-frailty model, despite their increased size. As in Table 14.1, the spatial frailty model has the best DIC value.

Tables 14.5 and 14.6 provide 2.5, 50, and 97.5 posterior percentiles for the main effects in our two frailty models, respectively. In both tables, all of the predictors are significant at the .05 level. Overall, the results are broadly similar to those from our earlier parametric analysis in Tables 14.2 and 14.3. For instance, the effect of being in the Native American group is again noteworthy. Under the CAR model, this covariate increases the posterior median hazard rate by a factor of $e^{0.599} = 1.82$. The benefit of fitting the spatial CAR structure is also seen again in the reduction of the length of the 95% credible intervals for the spatial model compared to the i.i.d. model. Most notably, the 95% credible set for the effect of "mother's age" is $(-0.054, -0.014)$ under the nonspatial frailty model (Table 14.5), but $(-0.042, -0.013)$ under the CAR frailty model (Table 14.6), a reduction in length of roughly 28%. Thus overall, adding spatial structure to the frailty terms appears to be reasonable and beneficial. Maps analogous to Figures 14.1 and 14.2 (not shown) reveal a very similar story. ■

### 14.3 Spatiotemporal models

In this section we follow Banerjee and Carlin (2003) to develop a semiparametric (Cox) hierarchical Bayesian frailty model for capturing spatiotemporal heterogeneity in survival data. We then use these models to describe the pattern of breast cancer in the 99 counties of Iowa while accounting for important covariates, spatially correlated differences in the hazards among the counties, and possible space-time interactions.

We begin by extending the framework of the preceding section to incorporate temporal dependence. Here we have $t_{ijk}$ as the response (time to death) for the $j$th subject residing in the $i$th county who was diagnosed in the $k$th year, while the individual-specific vector of covariates is now denoted by $\mathbf{x}_{ijk}$, for $i = 1, 2, ..., I$, $k = 1, \ldots, K$, and $j = 1, 2, ..., n_{ik}$. We

| Covariate | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Intercept | −1.961 | −1.532 | −0.845 |
| Sex (boys = 0) | | | |
| girls | −0.351 | −0.290 | −0.217 |
| Race (white = 0) | | | |
| black | −0.359 | −0.217 | −0.014 |
| Native American | 0.324 | 0.599 | 0.919 |
| unknown | 0.365 | 0.863 | 1.316 |
| Mother's age | −0.042 | −0.026 | −0.013 |
| Birth weight in kg | −1.325 | −1.301 | −1.283 |
| Total births | 0.088 | 0.135 | 0.193 |

Table 14.6 *Posterior summaries for the CAR semiparametric frailty model.*

note that "time" is now being used in two ways. The measurement or response is a survival time, but these responses are themselves observed at different areal units *and* different times (years). Furthermore, the spatial random effects $W_i$ in the preceding section are now modified to $W_{ik}$, to represent spatiotemporal frailties corresponding to the $i$th county for the $k$th diagnosis year. Our spatial frailty specification in (14.1) now becomes

$$h\left(t_{ijk};\mathbf{x}_{ijk}\right)=h_{0i}\left(t_{ijk}\right)\exp\left(\boldsymbol{\beta}^{T}\mathbf{x}_{ijk}+W_{ik}\right)\ . \tag{14.10}$$

Our CAR prior would now have conditional representation $W_{ik}|\ \ W_{(i'\neq i)k}\ \sim N(\overline{W}_{ik},1/(\lambda_k m_i))$.

Note that we can account for temporal correlation in the frailties by assuming that the $\lambda_k$ are themselves identically distributed from a common hyperprior (Subsection 11.7.1). A gamma prior (usually vague but proper) is often selected here, since this is particularly convenient for MCMC implementation. A flat prior for $\boldsymbol{\beta}$ is typically chosen, since this still admits a proper posterior distribution. Adaptive rejection (Gilks and Wild, 1992) or Metropolis-Hastings sampling are usually required to update the $\mathbf{W}_k$ and $\boldsymbol{\beta}$ parameters in a hybrid Gibbs sampler.

We remark that it would certainly be possible to include both spatial and nonspatial frailties, as mentioned in Subsection 14.1.1. This would mean supplementing our spatial frailties $W_{ik}$ with a collection of nonspatial frailties, say $V_{ik}\overset{iid}{\sim}N(0,1/\tau_k)$. We summarize our full hierarchical model as follows:

$$\begin{aligned}
L\left(\beta,\mathbf{W};\mathbf{t},\mathbf{x},\gamma\right) \ &\propto\ \prod_{k=1}^{K}\prod_{i=1}^{I}\prod_{j=1}^{n_{ik}}\left\{h_{0i}\left(t_{ijk};\mathbf{x}_{ijk}\right)\right\}^{\gamma_{ijk}} \\
&\quad\times\exp\left\{-H_{0i}\left(t_{ijk}\right)\exp\left(\boldsymbol{\beta}^{T}\mathbf{x}_{ijk}+W_{ik}+V_{ik}\right)\right\}, \\
\text{where}\ \ p(\mathbf{W}_k|\lambda_k) \ &\sim\ CAR(\lambda_k)\ \ p\left(\mathbf{V}_k|\tau_k\right)\ \sim\ N_I\left(\mathbf{0},\tau_k\mathbf{I}\right) \\
\text{and}\ \ \lambda_k \ &\sim\ G\left(a,b\right)\ \ ,\ \ \tau_k\ \sim\ G\left(c,d\right)\ \ \text{for}\ \ k=1,2,...,K\ .
\end{aligned}$$

In the sequel we adopt the beta mixture approach of Subsection 14.2.1 to model the baseline hazard functions $H_{0i}(t_{ijk})$ nonparametrically.

**Example 14.3** *(Analysis of Iowa SEER breast cancer data).* The National Cancer Institute's SEER program (`seer.cancer.gov`) is the most authoritative source of cancer data in the U.S., offering county-level summaries on a yearly basis for several states in various parts of the country. In particular, the database provides a cohort of 15,375 women in Iowa who were diagnosed with breast cancer starting in 1973, and have been undergoing treatment

| Covariate | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Age at diagnosis | 0.0135 | 0.0148 | 0.0163 |
| Number of primaries | –0.43 | –0.40 | –0.36 |
| Race (white = 0) | | | |
| black | –0.14 | 0.21 | 0.53 |
| other | –2.25 | –0.30 | 0.97 |
| Stage (local = 0) | | | |
| regional | 0.30 | 0.34 | 0.38 |
| distant | 1.45 | 1.51 | 1.58 |

Table 14.7 *Posterior summaries for the spatiotemporal frailty model.*

and have been progressively monitored since. Only those who have been identified as having died from metastasis of cancerous nodes in the breast are considered to have failed, while the rest (including those who might have died from metastasis of other types of cancer, or from other causes of death) are considered censored. By the end of 1998, 11,912 of the patients had died of breast cancer while the remaining were censored, either because they survived until the end of the study period, dropped out of the study, or died of causes other than breast cancer. For each individual, the data set records the time in months (1 to 312) that the patient survived, and her county of residence at diagnosis. Several individual-level covariates are also available, including race (black, white, or other), age at diagnosis, number of primaries (i.e., the number of other types of cancer diagnosed for this patient), and the stage of the disease (local, regional, or distant).

#### *14.3.0.1 Results for the full model*

We begin by summarizing our results for the spatiotemporal frailty model described above, i.e., the full model having both spatial frailties $W_{ik}$ and nonspatial frailties $V_{ik}$. We chose vague $G(0.01, 0.01)$ hyperpriors for the $\lambda_k$ and $\tau_k$ (having mean 1 but variance 100) in order to allow maximum flexibility in the partitioning of the frailties into spatial and nonspatial components. Best et al. (1999) suggest that a higher variance prior for the $\tau_k$ (say, a $G(0.001, 0.001)$) may lead to better prior "balance" between the spatial and nonspatial random effects, but there is controversy on this point and so we do not pursue it here. While overly diffuse priors (as measured for example as in Weiss, 1996) may result in weak identifiability of these parameters, their posteriors remain proper, and the impact of these priors on the posterior for the well-identified subset of parameters (including $\boldsymbol{\beta}$ and the log-relative hazards themselves) should be minimal (Daniels and Kass, 1999; Eberly and Carlin, 2000).

Table 14.7 provides 2.5, 50, and 97.5 posterior percentiles for the main effects (components of $\boldsymbol{\beta}$) in our model. All of the predictors *except* those having to do with race are significant at the .05 level. Since the reference group for the stage variable is local, we see that women with regional and distant (metastasized) diagnoses have higher and much higher hazard of death, respectively; the posterior median hazard rate increases by a factor of $e^{1.51} = 4.53$ for the latter group. Higher age at diagnosis also increases the hazard, but a larger number of primaries (the number of other types of cancer a patient is be suffering from) actually leads to a *lower* hazard, presumably due to the competing risk of dying from one of these other cancers.

Figure 14.5 maps the posterior medians of the frailties $W_{ik} + V_{ik}$ for the representative year 1986. We see clusters of counties with lower median frailties in the north-central and south-central parts of the state, and also clusters of counties with higher median frailties in the central, northeastern, and southeastern parts of the state.

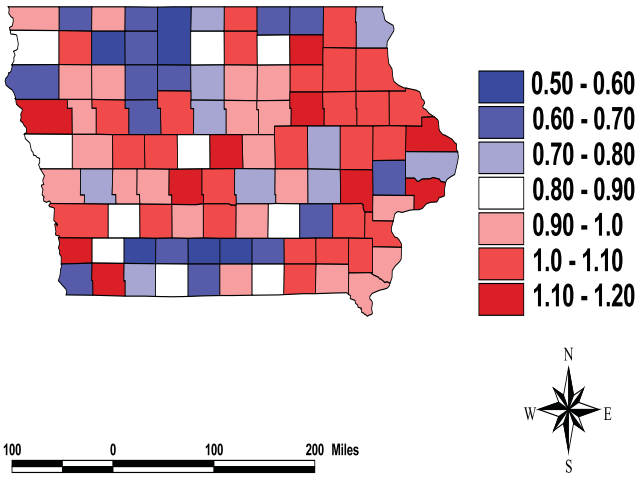


Figure 14.5 *Fitted spatiotemporal frailties, Iowa counties, 1986.*

Maps for other representative years showed very similar patterns, as well as an overall decreasing pattern in the frailties over time (see Banerjee and Carlin, 2003, for details). Figure 14.6 clarifies this pattern by showing boxplots of the posterior medians of the $W_{ik}$ over time (recall our full model does not have year-specific intercepts; the average of the $W_{ik}$ for year $k$ plays this role). We see an essentially horizontal trend during roughly the first half of our observation period, followed by a decreasing trend that seems to be accelerating. Overall the total decrease in median log hazard is about 0.7 units, or about a 50% reduction in hazard over the observation period. A cancer epidemiologist would likely be unsurprised by this decline, since it coincides with the recent rise in the use of mammography by American women.

#### *14.3.0.2 Bayesian model choice*

For model choice, we again turn to the DIC criterion. The first six lines of Table 14.8 provide $p_D$ and DIC values for our full model and several simplications thereof. Note the full model (sixth line) is estimated to have only just over 150 effective parameters, a substantial reduction (recall there are $2 \times 99 \times 26 = 5148$ random frailty parameters alone). Removing the spatial frailties $W_{ik}$ from the log-relative hazard has little impact on $p_D$, but substantial negative impact on the DIC score. By contrast, removing the nonspatial frailties $V_{ik}$ reduces (i.e., improves) both $p_D$ and DIC, consistent with our findings in the previous subsection. Further simplifying the model to having a single set of spatial frailties $W_i$ that do not vary with time (but now also reinserting year-specific intercepts $\alpha_k$) has little effect on $p_D$ but does improve DIC a bit more (though this improvement appears only slightly larger than the order of Monte Carlo error in our calculations). Even more drastic simplifications (eliminating the $W_i$, and perhaps even the $\alpha_k$) lead to further drops in $p_D$, but at the cost of unacceptably large increases in DIC. Thus our county-level breast cancer survival data seem to have strong spatial structure that is still unaccounted for by the covariates in Table 14.7, but structure that is fairly similar for all diagnosis years.

The last three lines of Table 14.8 reconsider the best three log-relative hazard models above, but where we now replace the semiparametric mixture baseline hazard with a Weibull hazard having region-specific baseline hazards $h_{0i}(t_{ijk}; \rho_i) = \rho_i t_{ijk}^{\rho_i - 1}$ (note the spatial frailties play the role of the second parameter customarily associated with the Weibull model). These fully parametric models offer small advantages in terms of parsimony (smaller $p_D$),

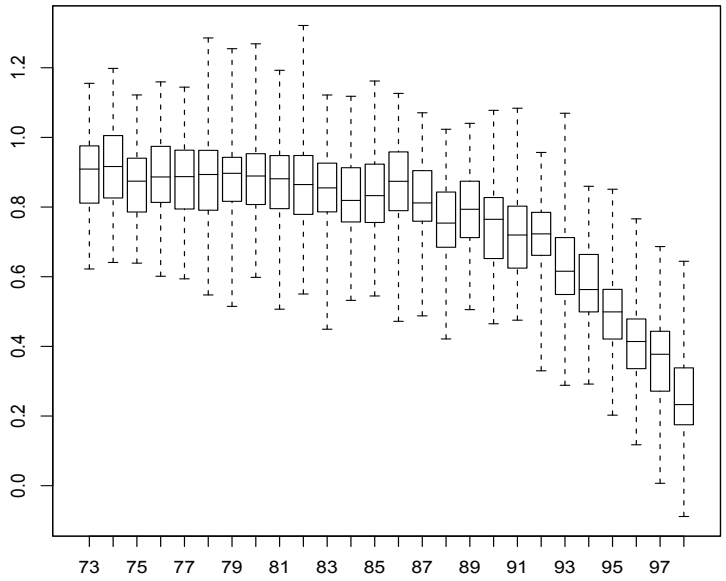


Figure 14.6 *Boxplots of posterior medians for the spatial frailties $W_{ik}$ over the Iowa counties for each year, k=1973, . . . , 1998.*

| Baseline hazard | Log-relative hazard | $p_D$ | DIC |
|---|---|---|---|
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk}$ | 6.17 | 780 |
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + \alpha_k$ | 33.16 | 743 |
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + \alpha_k + W_i$ | 80.02 | 187 |
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + W_{ik}$ | 81.13 | 208 |
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + V_{ik}$ | 149.45 | 732 |
| Semiparametric mixture | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + W_{ik} + V_{ik}$ | 151.62 | 280 |
| Weibull | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + \alpha_k + W_i$ | 79.22 | 221 |
| Weibull | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + W_{ik}$ | 80.75 | 239 |
| Weibull | $\boldsymbol{\beta}^T\mathbf{x}_{ijk} + W_{ik} + V_{ik}$ | 141.67 | 315 |

Table 14.8 *DIC and effective number of parameters $p_D$ for the competing models.*

but these gains are apparently more than outweighed by a corresponding degradation in fit (much larger DIC score). ■

### 14.4 Multivariate models ⋆

In this section we extend to multivariate spatial frailty modeling, using the MCAR model introduced in Subsection 10.1. In particular, we use a semiparametric model, and consider MCAR structure on both residual (spatial frailty) and regression (space-varying coefficient) terms. We also extend to the spatiotemporal case by including temporally correlated cohort effects (say, one for each year of initial disease diagnosis) that can be summarized and plotted over time. Example 14.4 illustrates the utility of our approach in an analysis of survival times of patients suffering from one or more types of cancer. We obtain posterior estimates of key

fixed effects, smoothed maps of both frailties and spatially varying coefficients, and compare models using the DIC criterion.

#### *14.4.0.3 Static spatial survival data with multiple causes of death*

Consider the following multivariate survival setting. Let $t_{ijk}$ denote the time to death or censoring for the $k$th patient having the $j$th type of primary cancer living in the $i$th county, $i = 1, \ldots, n$, $j = 1, \ldots, p$, $k = 1, \ldots, s_{ij}$, and let $\gamma_{ijk}$ be the corresponding death indicator. Let us write $\mathbf{x}_{ijk}$ as the vector of covariates for the above individual, and let $\mathbf{z}_{ijk}$ denote the vector of cancer indicators for this individual. That is, $\mathbf{z}_{ijk} = (z_{ijk1}, z_{ijk2}, ..., z_{ijkp})^T$ where $z_{ijkl} = 1$ if patient $ijk$ suffers from cancer type $l$, and 0 otherwise (note that $z_{ijkj} = 1$ by definition). Then we can write the likelihood of our proportional hazards model $L(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\Phi}; \mathbf{t}, \mathbf{x}, \boldsymbol{\gamma})$ as

$$\begin{array}{l} \prod_{i=1}^{n} \prod_{j=1}^{p} \prod_{k=1}^{s_{ij}} \left\{ h\left(t_{ijk}; \mathbf{x}_{ijk}, \mathbf{z}_{ijk}\right) \right\}^{\gamma_{ijk}} \\ \quad \times \exp\left\{ -H_{0i}\left(t_{ijk}\right) \exp\left( \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ijk}^T \boldsymbol{\theta} + \phi_{ij} \right) \right\}, \end{array} \tag{14.11}$$

where

$$h\left(t_{ijk}; \mathbf{x}_{ijk}, \mathbf{z}_{ijk}\right) = h_{0i}\left(t_{ijk}\right) \exp\left( \mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ijk}^T \boldsymbol{\theta} + \phi_{ij} \right) . \tag{14.12}$$

Here, $H_{0i}\left(t_{ijk}\right) = \int_0^{t_{ijk}} h_{0i}\left(u\right) du$, $\boldsymbol{\phi}_i = \left(\phi_{i1}, \phi_{i2}, ..., \phi_{in}\right)^T$, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are given flat priors, and

$$\boldsymbol{\Phi} \equiv \left( \boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_n^T \right)^T \sim MCAR\left(\rho, \boldsymbol{\Sigma}\right) ,$$

using the notation of Section 10.1. The region-specific baseline hazard functions $h_{0i}\left(t_{ijk}\right)$ are modeled using the beta mixture approach (Subsection 14.2.1) in such a way that the intercept in $\boldsymbol{\beta}$ remains estimable. We note that we could extend to a county *and* cancer-specific baseline hazard $h_{0ij}$; however, preliminary exploratory analyses of our data suggest such generality is not needed here.

Several alternatives to model formulation (14.12) immediately present themselves. For example, we could convert to a space-varying coefficients model (Assunção, 2003), replacing the log-relative hazard $\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ijk}^T \boldsymbol{\theta} + \phi_{ij}$ in (14.12) with

$$\mathbf{x}_{ijk}^T \boldsymbol{\beta} + \mathbf{z}_{ijk}^T \boldsymbol{\theta}_i , \tag{14.13}$$

where $\boldsymbol{\beta}$ again has a flat prior, but $\boldsymbol{\Theta} \equiv \left( \boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_n^T \right)^T \sim MCAR\left(\rho, \boldsymbol{\Sigma}\right)$. In Example 14.4 we apply this method to our cancer data set; we defer mention of still other log-relative hazard modeling possibilities until after this illustration.

#### *14.4.0.4 MCAR specification, simplification, and computing*

To efficiently implement the $MCAR\left(\rho, \boldsymbol{\Sigma}\right)$ as a prior distribution for our spatial process, suppose that we are using the usual 0-1 adjacency weights in $W$. Then recall from equation (10.4) that we may express the MCAR precision matrix $\mathbf{B} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\phi}}^{-1}$ in terms of the $n \times n$ adjacency matrix $W$ as

$$\mathbf{B} = \left(Diag(m_i) - \rho W\right) \otimes \Lambda ,$$

where we have added a propriety parameter $\rho$. Note that this is a Kronecker product of an $n \times n$ and a $p \times p$ matrix, thereby rendering $\mathbf{B}$ as $np \times np$ as required. In fact, $\mathbf{B}$ may be looked upon as the Kronecker product of two partial precision matrices: one for the spatial components, $\left(Diag(m_i) - \rho W\right)$ (depending upon their adjacency structure and number of neighbors), and another for the variation across diseases, given by $\Lambda$. We thus alter our notation slightly to $MCAR\left(\rho, \Lambda\right)$.

Also as a consequence of this form, a sufficient condition for positive definiteness of the dispersion matrix for this MCAR model becomes $|\rho| < 1$ (as in the univariate case). Negative smoothness parameters are not desirable, so we typically take $0 < \rho < 1$. We can now complete the Bayesian hierarchical formulation by placing appropriate priors on $\rho$ (say, a $Unif\,(0,1)$ or $Beta(18,2)$) and $\Lambda$ (say, a $Wishart\,(\rho, \Lambda_0)$).

The Gibbs sampler is the MCMC method of choice here, particularly because, as in the univariate case, it takes advantage of the MCAR's conditional specification. Adaptive rejection sampling may be used to sample the regression coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, while Metropolis steps with (possibly multivariate) Gaussian proposals may be employed for the spatial effects $\boldsymbol{\Phi}$. The full conditional for $\rho$ is nicely suited for slice sampling (see Subsection 5.3.3), given its bounded support. Finally, the full conditional for $\Lambda^{-1}$ emerges in closed form as an inverted Wishart distribution.

We conclude this subsection by recalling that our model can be generalized to admit different propriety parameters $\rho_j$ for different diseases (recall Section 10.1 and, in particular, the discussion surrounding (10.8)) and (10.9). We notate this model as $MCAR\,(\boldsymbol{\rho}, \Lambda)$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)^T$.

#### *14.4.0.5 Spatiotemporal survival data*

Here we extend our model to allow for cohort effects. Let $r$ index the year in which patient $ijk$ entered the study (i.e., the year in which the patient's primary cancer was diagnosed). Extending model (14.12) we obtain the log-relative hazard,

$$\mathbf{x}_{ijkr}^T \boldsymbol{\beta} + \mathbf{z}_{ijkr}^T \boldsymbol{\theta} + \phi_{ijr} \;, \tag{14.14}$$

with the obvious corresponding modifications to the likelihood (14.11). Here, $\boldsymbol{\phi}_{ir} = (\phi_{i1r}, \phi_{i2r}, ..., \phi_{ipr})^T$ and $\boldsymbol{\Phi}_r = \left(\boldsymbol{\phi}_{1r}^T, \ldots, \boldsymbol{\phi}_{nr}^T\right)^T \overset{ind}{\sim} MCAR\,(\rho_r, \Lambda_r)$. This permits addition of an exchangeable prior structure,

$$\rho_r \overset{iid}{\sim} Beta(a,b) \text{ and } \Lambda_r \overset{iid}{\sim} Wishart(\rho, \Lambda_0) \;,$$

where we may choose fixed values for $a, b, \rho$, and $\Lambda_0$, or place hyperpriors on them and estimate them from the data. Note also the obvious extension to disease-specific $\rho_{jr}$, as mentioned at the end of the previous subsection.

**Example 14.4** *(Application to Iowa SEER multiple cancer survival data).* We illustrate the approach with an analysis of SEER data on 17,146 patients from the 99 counties of the state of Iowa who have been diagnosed with cancer between 1992 and 1998, and who have a well-identified primary cancer. Our covariate vector $\mathbf{x}_{ijk}$ consists of a constant (intercept), a gender indicator, the age of the patient, indicators for race with "white" as the baseline, indicators for the stage of the primary cancer with "local" as the baseline, and indicators for year of primary cancer diagnosis (cohort) with the first year (1992) as the baseline. The vector $\mathbf{z}_{ijk}$ comprises the indicators of which cancers the patient has; the corresponding parameters will thus capture the effect of these cancers on the hazards regardless of whether they emerge as primary or secondary.

With regard to modeling details, we used five separate (cancer-specific) propriety parameters $\rho_j$ having an exchangeable $Beta(18,2)$ prior, and a vague $Wishart\,(\rho, \Lambda_0)$ for $\Lambda$, where $\rho = 5$ and $\Lambda_0 = 0.01 I_{5\times 5}$. (Results for $\boldsymbol{\beta}, \boldsymbol{\theta}$, and $\boldsymbol{\Phi}$ under a $U(0,1)$ prior for the $\rho_j$ were broadly similar.) Table 14.9 gives posterior summaries for the main effects $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$; note that $\boldsymbol{\theta}$ is estimable despite the presence of the intercept since many individuals have more than one cancer. No race or cohort effects emerged as significantly different from zero, so they have been deleted; all remaining effects are shown here. All of these effects are significant and in the directions one would expect. In particular, the five cancer effects

| Variable | 2.5% | 50% | 97.5% |
|---|---|---|---|
| Intercept | 0.102 | 0.265 | 0.421 |
| Sex (female = 0) | 0.097 | 0.136 | 0.182 |
| Age | 0.028 | 0.029 | 0.030 |
| Stage of primary cancer (local = 0) | | | |
| regional | 0.322 | 0.373 | 0.421 |
| distant | 1.527 | 1.580 | 1.654 |
| Type of primary cancer | | | |
| colorectal | 0.112 | 0.252 | 0.453 |
| gallbladder | 1.074 | 1.201 | 1.330 |
| pancreas | 1.603 | 1.701 | 1.807 |
| small intestine | 0.128 | 0.287 | 0.445 |
| stomach | 1.005 | 1.072 | 1.141 |

Table 14.9 *Posterior quantiles for the fixed effects in the MCAR frailty model.*

| Dallas County | Colo-rectal | Gall-bladder | Pancreas | Small intestine | Stomach |
|---|---|---|---|---|---|
| Colorectal | 0.852 | 0.262 | 0.294 | 0.413 | 0.464 |
| Gallbladder | | 1.151 | 0.314 | 0.187 | 0.175 |
| Pancreas | | | 0.846 | 0.454 | 0.528 |
| Small intestine | | | | 1.47 | 0.413 |
| Stomach | | | | | 0.908 |
| **Clay County** | **Colo-rectal** | **Gall-bladder** | **Pancreas** | **Small intestine** | **Stomach** |
| Colorectal | 0.903 | 0.215 | 0.273 | 0.342 | 0.352 |
| Gallbladder | | 1.196 | 0.274 | 0.128 | 0.150 |
| Pancreas | | | 0.852 | 0.322 | 0.402 |
| Small intestine | | | | 1.515 | 0.371 |
| Stomach | | | | | 1.068 |

Table 14.10 *Posterior variances and correlation summaries, Dallas and Clay Counties, MCAR spatial frailty model. Diagonal elements are estimated variances, while off-diagonal elements are estimated correlations.*

are consistent with results of previous modeling of this and similar data sets, with pancreatic cancer emerging as the most deadly (posterior median log relative hazard 1.701) and colorectal and small intestinal cancer relatively less so (.252 and .287, respectively).

Table 14.10 gives posterior variance and correlation summaries for the frailties $\phi_{ij}$ among the five cancers for two representative counties, Dallas (urban; Des Moines area) and Clay (rural northwest). Note that the correlations are as high as 0.528 (pancreas and stomach in Dallas County), suggesting the need for the multivariate structure inherent in our MCAR frailty model. Note also that summarizing the posterior distribution of $\Lambda^{-1}$ would be inappropriate here, since despite the Kronecker structure here, $\Lambda^{-1}$ cannot be directly interpreted as a primary cancer covariance matrix across counties.

Turning to geographic summaries, Figure 14.7 shows `ArcView` maps of the posterior means of the MCAR spatial frailties $\phi_{ij}$. Recall that in this model, the $\phi_{ij}$ play the role of spatial residuals, capturing any spatial variation not already accounted for by the spatial main effects $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The lack of spatial pattern in these maps suggest there is little additional spatial "story" in the data beyond what is already being told by the fixed effects. However, the map scales reveal that one cancer (gallbladder) is markedly different from the

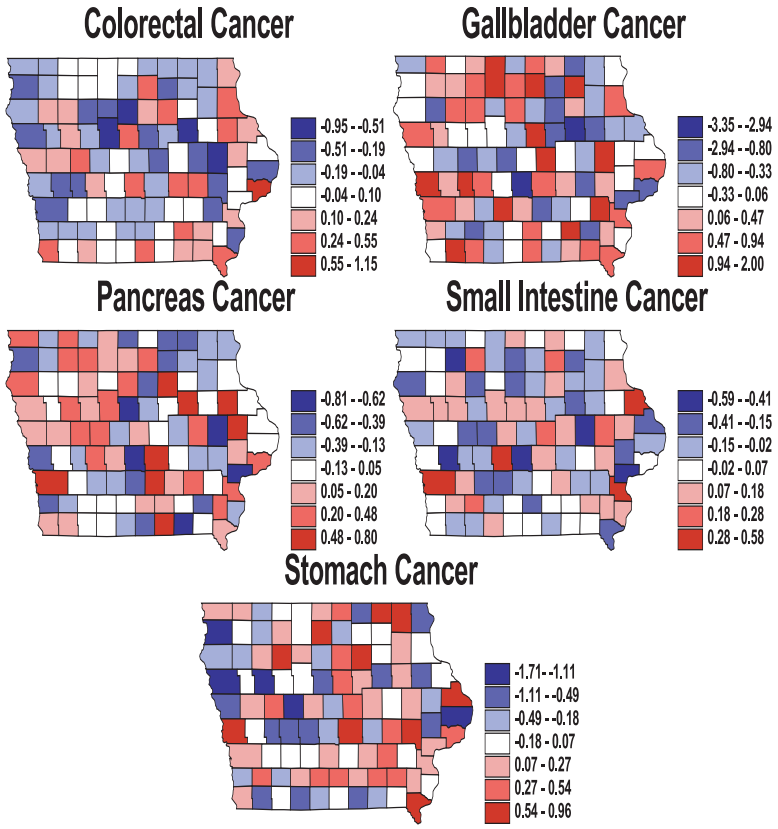


Figure 14.7 *Posterior mean spatial frailties, Iowa cancer data, static spatial MCAR model.*

others, both in terms of total range of the mean frailties (rather broad) and their center (negative; the other four are centered near 0).

Next, we change from the MCAR spatial frailty model to the MCAR spatially varying coefficients model (14.13). This model required a longer burn-in period (20,000 instead of 10,000), but otherwise our prior and MCMC control parameters remain unchanged. Figure 14.8 shows `ArcView` maps of the resulting posterior means of the spatially varying coefficients $\theta_{ij}$. Unlike the $\phi_{ij}$ in the previous model, these parameters are not "residuals," but the effects of the presence of the primary cancer indicated on the death rate in each county. Clearly these maps show a strong spatial pattern, with (for example) southwestern Iowa counties having relatively high fitted values for pancreatic and stomach cancer, while southeastern counties fare relatively poorly with respect to colorectal and small intestinal cancer. The overall levels for each cancer are consistent with those given for the corresponding fixed effects $\boldsymbol{\theta}$ in Table 14.9 for the spatial frailty model.

Table 14.11 gives the effective model sizes $p_D$ and DIC scores for a variety of spatial survival models. The first two listed (fixed effects only and standard CAR frailty) have few effective parameters, but also poor (large) DIC scores. The MCAR spatial frailty models (which place the MCAR on $\boldsymbol{\Phi}$) fare better, especially when we add the disease-specific $\rho_j$ (the model summarized in Table 14.9, Table 14.10, and Figure 14.7). However, adding heterogeneity effects $\epsilon_{ij}$ to this model adds essentially no extra effective parameters, and is actually harmful to the overall DIC score (since we are adding complexity for little or no

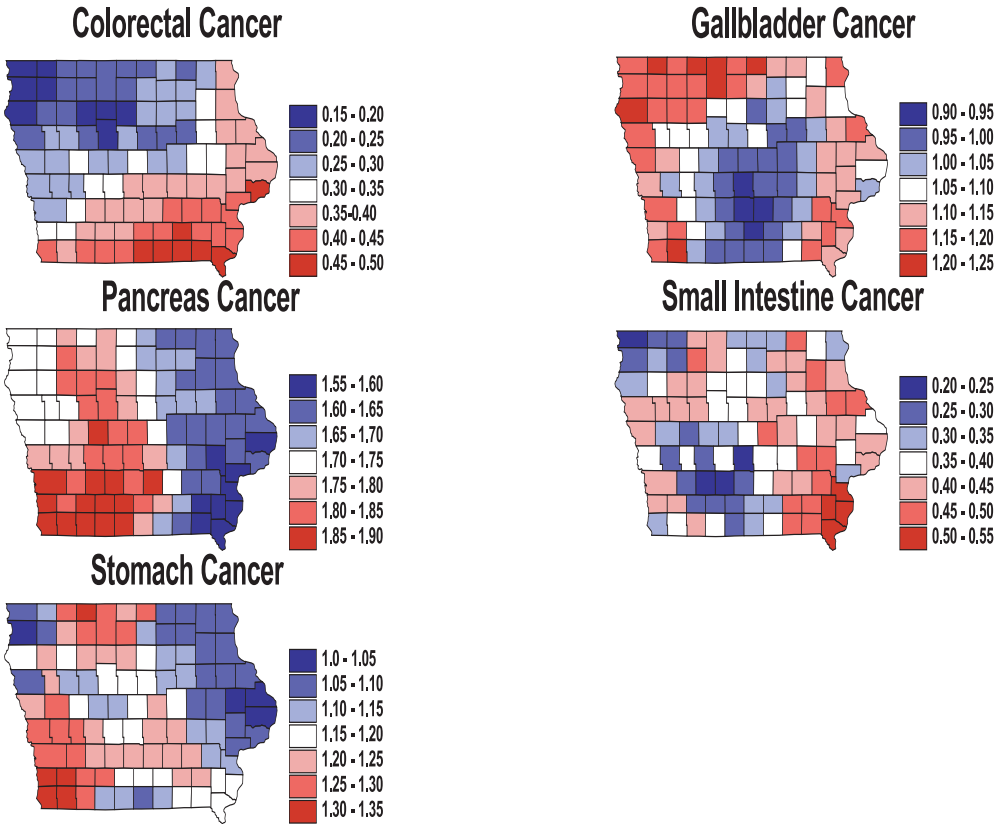


Figure 14.8 *Posterior mean spatially varying coefficients, Iowa cancer data, static spatial MCAR model.*

| Log-relative hazard model | $p_D$ | DIC |
|---|---|---|
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta}$ | 10.97 | 642 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta} + \phi_i,\ \boldsymbol{\phi} \sim CAR(\rho, \lambda)$ | 103.95 | 358 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta} + \phi_{ij},\ \boldsymbol{\Phi} \sim MCAR(\rho = 1, \Lambda)$ | 172.75 | 247 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta} + \phi_{ij},\ \boldsymbol{\Phi} \sim MCAR(\rho, \Lambda)$ | 172.40 | 246 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta} + \phi_{ij},\ \boldsymbol{\Phi} \sim MCAR(\rho_1, \ldots, \rho_5, \Lambda)$ | 175.71 | 237 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta} + \phi_{ij} + \epsilon_{ij},\ \boldsymbol{\Phi} \sim MCAR(\rho_1, \ldots, \rho_5, \Lambda),$ $\epsilon_{ij} \overset{iid}{\sim} N(0, \tau^2)$ | 177.25 | 255 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta}_i,\ \boldsymbol{\Theta} \sim MCAR(\rho, \Lambda)$ | 169.42 | 235 |
| $\mathbf{x}_{ijk}^T\boldsymbol{\beta} + \mathbf{z}_{ijk}^T\boldsymbol{\theta}_i,\ \boldsymbol{\Theta} \sim MCAR(\rho_1, \ldots, \rho_5, \Lambda)$ | 171.46 | 229 |

Table 14.11 *DIC comparison, spatial survival models for the Iowa cancer data.*

benefit in terms of fit). Finally, the two spatially varying coefficients models enjoy the best (smallest) DIC scores, but only by a small margin over the best spatial frailty model.

Finally, we fit the spatiotemporal extension (14.14) of our MCAR frailty model to the data where the cohort effect (year of study entry $r$) is taken into account. Year-by-year boxplots of the posterior median frailties (Figure 14.9) reveal the expected steadily decreasing trend for all five cancers, though it is not clear how much of this decrease is simply an artifact

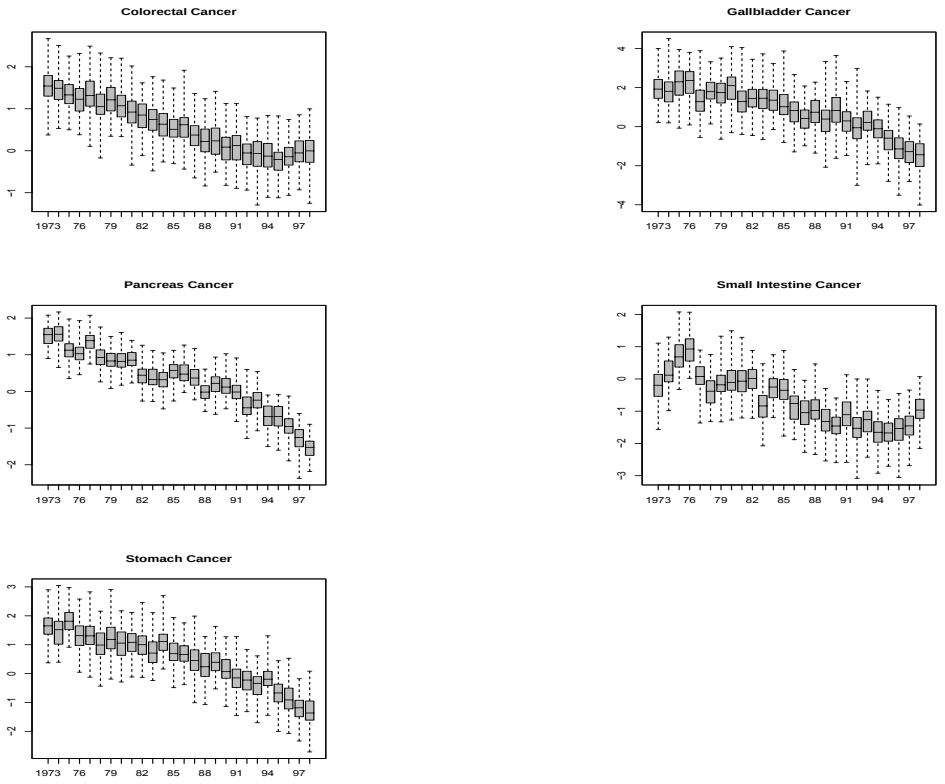


Figure 14.9 *Boxplots of posterior medians for the spatial frailties $\phi_{ijr}$ over the 99 Iowa counties for each year, $r = 1973, \ldots, 1998$.*

of the censoring of survival times for patients in more recent cohorts. The spatiotemporal extension of the spatially varying coefficients model (14.13) (i.e., $\mathbf{x}_{ijkr}^T\boldsymbol{\beta} + \mathbf{z}_{ijkr}^T\boldsymbol{\theta}_{ir}$) might well produce results that are temporally more interesting in this case. Incorporating change points, cancer start date measurement errors, and other model enhancements (say, interval censoring) might also be practically important model enhancements here. ■

## 14.5 Spatial cure rate models ⋆

In Section 14.1 we investigated spatially correlated frailties in traditional parametric survival models, choosing a random effects distribution to reflect the spatial structure in the problem. Sections 14.2 and 14.3 extended this approach to spatial and spatiotemporal settings within a semiparametric model.

In this section our ultimate goal is the proper analysis of a geographically referenced smoking cessation study, in which we observe subjects periodically through time to check for relapse following an initial quit attempt. Each patient is observed once each year for five consecutive years, whereupon the current average number of cigarettes smoked at each visit is recorded, along with the zip code of residence and several other potential explanatory variables. This data set requires us to extend the work of Carlin and Hodges (1999) in a number of ways. The primary extension involves the incorporation of a *cure fraction* in our models. In investigating the effectiveness of quitting programs, data typically reveal

many former smokers having successfully given up smoking, and as such may be thought of as "cured" of the deleterious habit. Incorporating such cure fractions in survival models leads to *cure rate models*, which are often applied in survival settings where the endpoint is a particular disease (say, breast cancer) which the subject may never reexperience. These models have a long history in the biostatistical literature, with the most popular perhaps being that of Berkson and Gage (1952). This model has been extensively studied in the statistical literature by a number of authors, including Farewell (1982, 1986), Goldman (1984), and Ewell and Ibrahim (1997). Recently, cure rates have been studied in somewhat more general settings by Chen, Ibrahim, and Sinha (1999) following earlier work by Yakovlev and Tsodikov (1996).

In addition, while this design can be analyzed as an ordinary right-censored survival model (with relapse to smoking as the endpoint), the data are perhaps more accurately viewed as *interval-censored*, since we actually observe only approximately annual intervals within which a failed quitter resumed smoking. We will consider both right- and interval-censored models, where in the former case we simply approximate the time of relapse by the midpoint of the corresponding time interval. Finally, we capture spatial variation through zip code-specific spatial random effects in the cure fraction or the hazard function, which in either case may act as spatial frailties. We find that incorporating the covariates and frailties into the hazard function is most natural (both intuitively and methodologically), especially after adopting a Weibull form for the baseline hazard.

### *14.5.1 Models for right- and interval-censored data*

#### *14.5.1.1 Right-censored data*

Our cure rate models are based on those of Chen et al. (1999) and derived assuming that some latent biological process is generating the observed data. Suppose there are $I$ regions and $n_i$ patients in the $i$th region. We denote by $T_{ij}$ the random variable for time to event (relapse, in our case) of the $j$th person in the $i$th region, where $j = 1, 2, ..., n_i$ and $i = 1, 2, ..., I$. (While acknowledging the presence of the regions in our notation, we postpone explicit spatial modeling to the next section.) Suppose that the $(i, j)$th individual has $N_{ij}$ potential latent (unobserved) risk factors, the presence of any of which (i.e., $N_{ij} \geq 1$) will ultimately lead to the event. For example, in cancer settings these factors may correspond to metastasis-competent tumor cells within the individual. Typically, there will be a number of subjects who do not undergo the event during the observation period, and are therefore considered censored. Thus, letting $U_{ijk}$, $k = 1, 2, ..., N_{ij}$ be the time to an event arising from the $k$th latent factor for the $(i, j)$th individual, the observed time to event for an uncensored individual is generated by $T_{ij} = \min\{U_{ijk},\, k = 1, 2, ..., N_{ij}\}$. If the $(i, j)$th individual is right-censored at time $t_{ij}$, none of the latent factors have led to an event by that time, and clearly $T_{ij} > t_{ij}$ (and in fact $T_{ij} = \infty$ if $N_{ij} = 0$).

Given $N_{ij}$, the $U_{ijk}$'s are independent with survival function $S(t|\Psi_{ij})$ and corresponding density function $f(t|\Psi_{ij})$. The parameter $\Psi_{ij}$ is a collection of all the parameters (including possible regression parameters) that may be involved in a parametric specification for the survival function $S$. In this section we will work with a two-parameter Weibull distribution specification for the density function $f(t|\Psi_{ij})$, where we allow the Weibull scale parameter $\rho$ to vary across the regions, and $\eta$, which may serve as a link to covariates in a regression setup, to vary across individuals. Therefore $f(t|\rho_i, \eta_{ij}) = \rho_i t^{\rho_i - 1} \exp(\eta_{ij} - t^{\rho_i} \exp(\eta_{ij}))$.

In terms of the hazard function $h$, $f(t|\rho_i, \eta_{ij}) = h(t|\rho_i, \eta_{ij}) S(t|\rho_i, \eta_{ij})$, with $h(t; \rho_i, \eta_{ij}) = \rho_i t^{\rho_i - 1} \exp(\eta_{ij})$ and $S(t|\rho_i, \eta_{ij}) = \exp(-t^{\rho_i} \exp(\eta_{ij}))$. Note we implicitly assume proportional hazards, with baseline hazard function $h_0(t|\rho_i) = \rho_i t^{\rho_i - 1}$. Thus an individual $ij$ who is censored at time $t_{ij}$ before undergoing the event contributes $(S(t_{ij}|\rho_i, \eta_{ij}))^{N_{ij}}$ to the likelihood, while an individual who experiences the event at time

$t_{ij}$ contributes $N_{ij}\left(S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{N_{ij}-1} f\left(t_{ij}|\rho_i,\eta_{ij}\right)$. The latter expression follows from the fact that the event is experienced when any one of the latent factors occurs. Letting $\nu_{ij}$ be the observed event indicator for individual $ij$, this person contributes

$$\begin{aligned} &L\left(t_{ij}|N_{ij},\rho_i,\eta_{ij},\nu_{ij}\right) \\ &\quad = \left(S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{N_{ij}(1-\nu_{ij})}\left(N_{ij}S\left(t_{ij}|\rho_i,\eta_{ij}\right)^{N_{ij}-1} f\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{\nu_{ij}} , \end{aligned}$$

and the joint likelihood for all the patients can now be expressed as

$$\begin{aligned} &L\left(\{t_{ij}\}\,|\,\{N_{ij}\},\{\rho_i\},\{\eta_{ij}\},\{\nu_{ij}\}\right) \\ &\quad = \prod_{i=1}^{I}\prod_{j=1}^{n_i} L\left(t_{ij}|N_{ij},\rho_i,\eta_{ij},\nu_{ij}\right) \\ &\quad = \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{N_{ij}(1-\nu_{ij})} \\ &\qquad\qquad \times\left(N_{ij}S\left(t_{ij}\rho_i,\eta_{ij}\right)^{N_{ij}-1} f\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{\nu_{ij}} \\ &\quad = \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{N_{ij}-\nu_{ij}}\left(N_{ij}f\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{\nu_{ij}} . \end{aligned}$$

This expression can be rewritten in terms of the hazard function as

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{N_{ij}}\left(N_{ij}h\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)^{\nu_{ij}} . \tag{14.15}$$

A Bayesian hierarchical formulation is completed by introducing prior distributions on the parameters. We will specify independent prior distributions $p\left(N_{ij}|\theta_{ij}\right)$, $p\left(\rho_i|\psi_\rho\right)$ and $p\left(\eta_{ij}|\psi_\eta\right)$ for $\{N_{ij}\}$, $\{\rho_i\}$, and $\{\eta_{ij}\}$, respectively. Here, $\psi_\rho$, $\psi_\eta$, and $\{\theta_{ij}\}$ are appropriate hyperparameters. Assigning independent hyperpriors $p\left(\theta_{ij}|\psi_\theta\right)$ for $\{\theta_{ij}\}$ and assuming the hyperparameters $\psi=(\psi_\rho,\psi_\eta,\psi_\theta)$ to be fixed, the posterior distribution for the parameters, $p\left(\{\theta_{ij}\},\{\eta_{ij}\},\{N_{ij}\},\{\rho_i\}\,|\,\{t_{ij}\},\{\nu_{ij}\}\right)$, is easily found (up to a proportionality constant) using (14.15) as

$$\begin{aligned} &\prod_{i=1}^{I}\Big\{p\left(\rho_i|\psi_\rho\right)\prod_{j=1}^{n_i}\left[S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right]^{N_{ij}}\left[N_{ij}h\left(t_{ij}|\rho_i,\eta_{ij}\right)\right]^{\nu_{ij}} \\ &\qquad \times p\left(N_{ij}|\theta_{ij}\right)p\left(\eta_{ij}|\psi_\eta\right)p\left(\theta_{ij}|\psi_\theta\right)\Big\} . \end{aligned}$$

Chen et al. (1999) assume that the $N_{ij}$ are distributed as independent Poisson random variables with mean $\theta_{ij}$, i.e., $p\left(N_{ij}|\theta_{ij}\right)$ is $Poisson\left(\theta_{ij}\right)$. In this setting it is easily seen that the survival distribution for the $(i,j)$th patient, $P\left(T_{ij}\geq t_{ij}|\rho_i,\eta_{ij}\right)$, is given by $\exp\left\{-\theta_{ij}\left(1-S\left(t_{ij}|\rho_i,\eta_{ij}\right)\right)\right\}$. Since $S\left(t_{ij}|\rho_i,\eta_{ij}\right)$ is a proper survival function (corresponding to the latent factor times $U_{ijk}$), as $t_{ij}\to\infty$, $P\left(T_{ij}\geq t_{ij}|\rho_i,\eta_{ij}\right)\to\exp\left(-\theta_{ij}\right)>0$. Thus we have a subdistribution for $T_{ij}$ with a *cure fraction* given by $\exp\left(-\theta_{ij}\right)$. Here a hyperprior on the $\theta_{ij}$'s would have support on the positive real line.

While there could certainly be multiple latent factors that increase the risk of smoking relapse (age started smoking, occupation, amount of time spent driving, tendency toward addictive behavior, etc.), this is rather speculative and certainly not as justifiable as in the cancer setting for which the multiple factor approach was developed (where $N_{ij}>1$ is biologically motivated). As such, we instead form our model using a single, omnibus, "propensity for relapse" latent factor. In this case, we think of $N_{ij}$ as a *binary* variable, and specify $p\left(N_{ij}|\theta_{ij}\right)$ as Bernoulli $(1-\theta_{ij})$. In this setting it is easier to look at the survival distribution after marginalizing out the $N_{ij}$. In particular, note that

$$P\left(T_{ij}\geq t_{ij}|\rho_i,\eta_{ij},N_{ij}\right)=\begin{cases} S\left(t_{ij}|\rho_i,\eta_{ij}\right) , & N_{ij}=1 \\ 1 , & N_{ij}=0 \end{cases} .$$

That is, if the latent factor is absent, the subject is cured (does not experience the event).

Marginalizing over the Bernoulli distribution for $N_{ij}$, we obtain for the $(i,j)$th patient the survival function $S^*\left(t_{ij}|\theta_{ij},\rho_i,\eta_{ij}\right) \equiv P\left(T_{ij} \geq t_{ij}|\rho_i,\eta_{ij}\right) = \theta_{ij} + (1-\theta_{ij})\, S\left(t_{ij}|\rho_i,\eta_{ij}\right)$, which is the classic cure-rate model attributed to Berkson and Gage (1952) with cure fraction $\theta_{ij}$. Now we can write the likelihood function for the data marginalized over $\{N_{ij}\}$, $L\left(\{t_{ij}\} \mid \{\rho_i\},\{\theta_{ij}\},\{\eta_{ij}\},\{\nu_{ij}\}\right)$, as

$$\begin{aligned}&\textstyle\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left[S^*\left(t_{ij}|\theta_{ij},\rho_i,\eta_{ij}\right)\right]^{1-\nu_{ij}}\left(-\frac{d}{dt_{ij}}S^*\left(t_{ij}|\theta_{ij},\rho_i,\eta_{ij}\right)\right)^{\nu_{ij}}\\ &\quad=\textstyle\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left[S^*\left(t_{ij}|\theta_{ij},\rho_i,\eta_{ij}\right)\right]^{1-\nu_{ij}}\left[(1-\theta_{ij})\, f\left(t_{ij}|\rho_i,\eta_{ij}\right)\right]^{\nu_{ij}}\ ,\end{aligned}$$

which in terms of the hazard function becomes

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i}\left[S^*\left(t_{ij}|\theta_{ij},\rho_i,\eta_{ij}\right)\right]^{1-\nu_{ij}}\left[(1-\theta_{ij})\, S\left(t_{ij}|\rho_i,\eta_{ij}\right) h\left(t_{ij}|\rho_i,\eta_{ij}\right)\right]^{\nu_{ij}}\ , \tag{14.16}$$

where the hyperprior for $\theta_{ij}$ has support on $(0,1)$. Now the posterior distribution of the parameters is proportional to

$$L\left(\{t_{ij}\} \mid \{\rho_i\},\{\theta_{ij}\},\{\eta_{ij}\},\{\nu_{ij}\}\right)\prod_{i=1}^{I}\left\{p\left(\rho_i|\psi_\rho\right)\prod_{j=1}^{n_i}p\left(\eta_{ij}|\psi_\eta\right)p\left(\theta_{ij}|\psi_\theta\right)\right\}\ . \tag{14.17}$$

Turning to the issue of incorporating covariates, in the general setting with $N_{ij}$ assumed to be distributed Poisson, Chen et al. (1999) propose their introduction in the cure fraction through a suitable link function $g$, so that $\theta_{ij} = g\left(\mathbf{x}_{ij}^T\widetilde{\boldsymbol{\beta}}\right)$, where $g$ maps the entire real line to the positive axis. This is sensible when we believe that the risk factors affect the probability of an individual being cured. Proper posteriors arise for the regression coefficients $\widetilde{\boldsymbol{\beta}}$ even under improper priors. Unfortunately, this is no longer true when $N_{ij}$ is Bernoulli (i.e., in the Berkson and Gage model). Vague but proper priors may still be used, but this makes the parameters difficult to interpret, and can often lead to poor MCMC convergence.

Since a binary $N_{ij}$ seems most natural in our setting, we instead introduce covariates into $S\left(t_{ij}|\rho_i,\eta_{ij}\right)$ through the Weibull link $\eta_{ij}$, i.e., we let $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta}$. This seems intuitively more reasonable anyway, since now the covariates influence the underlying factor that brings about the smoking relapse (and thus the rapidity of this event). Also, proper posteriors arise here for $\boldsymbol{\beta}$ under improper posteriors even though $N_{ij}$ is binary. As such, henceforth we will only consider the situation where the covariates enter the model in this way (through the Weibull link function). This means we are unable to separately estimate the effect of the covariates on both the *rate* of relapse and the *ultimate level* of relapse, but "fair" estimation here (i.e., allocating the proper proportions of the covariates' effects to each component) is not clear anyway since flat priors could be selected for $\boldsymbol{\beta}$, but not for $\widetilde{\boldsymbol{\beta}}$. Finally, all of our subsequent models also assume a constant cure fraction for the entire population (i.e., we set $\theta_{ij} = \theta$ for all $i,j$).

Note that the posterior distribution in (14.17) is easily modified to incorporate covariates. For example, with $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta}$, we replace $\prod_{ij} p\left(\eta_{ij}|\psi_\eta\right)$ in (14.17) with $p\left(\boldsymbol{\beta}|\psi_\beta\right)$, with $\psi_\beta$ as a fixed hyperparameter. Typically a flat or vague Gaussian prior may be taken for $p\left(\boldsymbol{\beta}|\psi_\beta\right)$.

#### *14.5.1.2 Interval-censored data*

The formulation above assumes that our observed data are right-censored. This means that we are able to observe the actual relapse time $t_{ij}$ when it occurs prior to the final office visit. In reality, our study (like many others of its kind) is only able to determine patient status at the office visits themselves, meaning we observe only a time *interval* $(t_{ijL}, t_{ijU})$ within

which the event (in our case, smoking relapse) is known to have occurred. For patients who did not resume smoking prior to the end of the study we have $t_{ijU} = \infty$, returning us to the case of right-censoring at time point $t_{ijL}$. Thus we now set $\nu_{ij} = 1$ if subject $ij$ is interval-censored (i.e., experienced the event), and $\nu_{ij} = 0$ if the subject is right-censored.

Following Finkelstein (1986), the general interval-censored cure rate likelihood, $L\left(\{(t_{ijL}, t_{ijU})\} \mid \{N_{ij}\}, \{\rho_i\}, \{\eta_{ij}\}, \{\nu_{ij}\}\right)$, is given by

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i} \left[S\left(t_{ijL}|\rho_i,\eta_{ij}\right)\right]^{N_{ij}-\nu_{ij}} \left\{N_{ij}\left[S\left(t_{ijL}|\rho_i,\eta_{ij}\right) - S\left(t_{ijU}|\rho_i,\eta_{ij}\right)\right]\right\}^{\nu_{ij}}$$
$$= \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left[S\left(t_{ijL}|\rho_i,\eta_{ij}\right)\right]^{N_{ij}} \left\{N_{ij}\left(1 - \frac{S\left(t_{ijU}|\rho_i,\eta_{ij}\right)}{S\left(t_{ijL}|\rho_i,\eta_{ij}\right)}\right)\right\}^{\nu_{ij}}.$$

As in the previous section, in the Bernoulli setup after marginalizing out the $\{N_{ij}\}$ the foregoing becomes $L\left(\{(t_{ijL}, t_{ijU})\} \mid \{\rho_i\}, \{\theta_{ij}\}, \{\eta_{ij}\}, \{\nu_{ij}\}\right)$, and can be written as

$$\prod_{i=1}^{I}\prod_{j=1}^{n_i} S^*\left(t_{ijL}|\theta_{ij},\rho_i,\eta_{ij}\right)\left\{1 - \frac{S^*\left(t_{ijU}|\theta_{ij},\rho_i,\eta_{ij}\right)}{S^*\left(t_{ijL}|\theta_{ij},\rho_i,\eta_{ij}\right)}\right\}^{\nu_{ij}}. \tag{14.18}$$

We omit details (similar to those in the previous section) arising from the Weibull parametrization and subsequent incorporation of covariates through the link function $\eta_{ij}$.

### *14.5.2 Spatial frailties in cure rate models*

The development of the hierarchical framework in the preceding section acknowledged the data as coming from $I$ different geographical regions (clusters). Such clustered data are common in survival analysis and often modeled using cluster-specific frailties $\phi_i$. As with the covariates, we will introduce the frailties $\phi_i$ through the Weibull link as intercept terms in the log-relative risk; that is, we set $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i$.

Here we allow the $\phi_i$ to be spatially correlated across the regions; similarly we would like to permit the Weibull baseline hazard parameters, $\rho_i$, to be spatially correlated. A natural approach in both cases is to use a univariate CAR prior. While one may certainly employ separate, independent CAR priors on $\boldsymbol{\phi}$ and $\boldsymbol{\zeta} \equiv \{\log\rho_i\}$, another option is to allow these two spatial priors to themselves be correlated. In other words, we may want a bivariate spatial model for the $\delta_i = (\phi_i, \zeta_i)^T = (\phi_i, \log\rho_i)^T$. As mentioned in Sections 10.1 and 14.4, we may use the MCAR distribution for this purpose. In our setting, the MCAR distribution on the concatenated vector $\boldsymbol{\delta} = (\boldsymbol{\phi}^T, \boldsymbol{\zeta}^T)^T$ is Gaussian with mean $\mathbf{0}$ and precision matrix $\Lambda^{-1} \otimes (Diag\,(m_i) - \rho W)$, where $\Lambda$ is a $2 \times 2$ symmetric and positive definite matrix, $\rho \in (0, 1)$, and $m_i$ and $W$ remain as above. In the current context, we may also wish to allow different smoothness parameters (say, $\rho_1$ and $\rho_2$) for $\boldsymbol{\phi}$ and $\boldsymbol{\zeta}$, respectively, as in Section 14.4. Henceforth, in this section we will denote the proper MCAR with a common smoothness parameter by $MCAR\,(\rho, \Lambda)$, and the multiple smoothness parameter generalized MCAR by $MCAR\,(\rho_1, \rho_2, \Lambda)$. Combined with independent (univariate) CAR models for $\boldsymbol{\phi}$ and $\boldsymbol{\zeta}$, these offer a broad range of potential spatial models.

### *14.5.3 Model comparison*

Suppose we let $\Omega$ denote the set of all model parameters, so that the deviance statistic (5.10) becomes

$$D(\Omega) = -2\log f(\mathbf{y}|\Omega) + 2\log h(\mathbf{y})\,. \tag{14.19}$$

When DIC is used to compare nested models in standard exponential family settings, the unnormalized likelihood $L(\Omega; \mathbf{y})$ is often used in place of the normalized form $f(\mathbf{y}|\Omega)$ in

(14.19), since in this case the normalizing function $m(\Omega) = \int L(\Omega; \mathbf{y}) d\mathbf{y}$ will be free of $\Omega$ and constant across models, hence contribute equally to the DIC scores of each (and thus have no impact on model selection). However, in settings where we require comparisons across different likelihood distributional forms, it appears one must be careful to use the properly scaled joint density $f(\mathbf{y}|\Omega)$ for each model.

We argue that use of the usual proportional hazards likelihood (which of course is not a joint density function) *is* in fact appropriate for DIC computation here, provided we make a fairly standard assumption regarding the relationship between the survival and censoring mechanisms generating the data. Specifically, suppose the distribution of the censoring times is independent of that of the survival times *and* does not depend upon the survival model parameters (i.e., independent, noninformative censoring). Let $g(t_{ij})$ denote the density of the censoring time for the $ij$th individual, with corresponding survival (1-cdf) function $R(t_{ij})$. Then the right-censored likelihood (14.16) can be extended to the joint likelihood specification,

$$\prod_{i=1}^{I} \prod_{j=1}^{n_i} \left[S^*\left(t_{ij}|\theta_{ij}, \rho_i, \eta_{ij}\right)\right]^{1-\nu_{ij}} \times \left[\left(1-\theta_{ij}\right) S\left(t_{ij}|\rho_i, \eta_{ij}\right) h\left(t_{ij}|\rho_i, \eta_{ij}\right)\right]^{\nu_{ij}} \left[R\left(t_{ij}\right)\right]^{\nu_{ij}} \left[g\left(t_{ij}\right)\right]^{1-\nu_{ij}} ,$$

as for example in Le (1997, pp. 69–70). While not a joint probability density, this likelihood is still an everywhere nonnegative and integrable function of the survival model parameters $\Omega$, and thus suitable for use with the Kullback-Leibler divergences that underlie DIC (Spiegelhalter et al., 2002, p. 586). But by assumption, $R(t)$ and $g(t)$ do not depend upon $\Omega$. Thus, like an $m(\Omega)$ that is free of $\Omega$, they may be safely ignored in both the $p_D$ and DIC calculations. Note this same argument implies that we can use the unnormalized likelihood (14.16) when comparing not only nonnested parametric survival models (say, Weibull versus gamma), but even parametric and semiparametric models (say, Weibull versus Cox) provided our definition of "likelihood" is comparable across models.

Note also that here our "focus" (in the nomenclature of Spiegelhalter et al., 2002) is solely on $\Omega$. An alternative would be instead to use a missing data formulation, where we include the likelihood contribution of $\{s_{ij}\}$, the collection of latent survival times for the right-censored individuals. Values for both $\Omega$ and the $\{s_{ij}\}$ could then be imputed along the lines given by Cox and Oakes (1984, pp. 165–166) for the EM algorithm or Spiegelhalter et al. (1995b, the "mice" example) for the Gibbs sampler. This would alter our focus from $\Omega$ to $(\Omega, \{s_{ij}\})$, and $p_D$ would reflect the correspondingly larger effective parameter count.

Turning to the interval censored case, here matters are only a bit more complicated. Converting the interval-censored likelihood (14.18) to a joint likelihood specification yields

$$\prod_{i=1}^{I} \prod_{j=1}^{n_i} S^*\left(t_{ijL}|\theta_{ij}, \rho_i, \eta_{ij}\right) \left(1 - \frac{S^*\left(t_{ijU}|\theta_{ij}, \rho_i, \eta_{ij}\right)}{S^*\left(t_{ijL}|\theta_{ij}, \rho_i, \eta_{ij}\right)}\right)^{\nu_{ij}} \times \left[R\left(t_{ijL}\right)\right]^{\nu_{ij}} \left(1 - \frac{R\left(t_{ijU}\right)}{R\left(t_{ijL}\right)}\right)^{\nu_{ij}} \left[g\left(t_{ijL}\right)\right]^{1-\nu_{ij}} .$$

Now $\left[R\left(t_{ijL}\right)\right]^{\nu_{ij}} \left(1 - R\left(t_{ijU}\right)/R\left(t_{ijL}\right)\right)^{\nu_{ij}} \left[g\left(t_{ijL}\right)\right]^{1-\nu_{ij}}$ is the function absorbed into $m(\Omega)$, and is again free of $\Omega$. Thus again, use of the usual form of the interval-censored likelihood presents no problems when comparing models within the interval-censored framework (including nonnested parametric models, or even parametric and semiparametric models).

Note that it does *not* make sense to compare a particular right-censored model with a particular interval-censored model. The form of the available data is different; model comparison is only appropriate to a given data set.

**Example 14.5** *(Smoking cessation data).* We illustrate our methods using the aforementioned study of smoking cessation, a subject of particular interest in studies of lung health

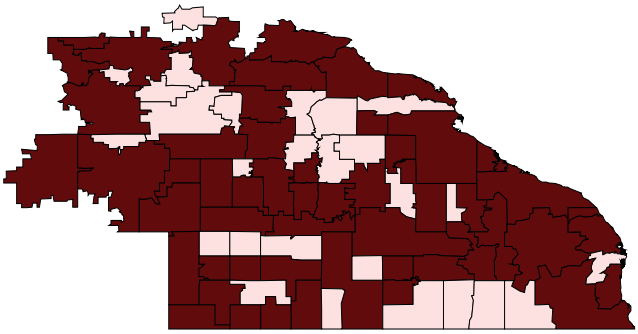

Figure 14.10 *Map showing missingness pattern for the smoking cessation data: lightly shaded regions are those having no responses.*

| Model | Log-relative risk | $pD$ | DIC |
|---|---|---|---|
| 1 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ \phi_i \overset{iid}{\sim} N(0, \tau_\phi),\ \rho_i = \rho\ \forall\ i$ | 10.3 | 438 |
| 2 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ \{\phi_i\} \sim CAR(\lambda_\phi),\ \rho_i = \rho\ \forall\ i$ | 9.4 | 435 |
| 3 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ \phi_i \overset{iid}{\sim} N(0, \tau_\phi),\ \zeta_i \overset{iid}{\sim} N(0, \tau_\zeta)$ | 13.1 | 440 |
| 4 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ \{\phi_i\} \sim CAR(\lambda_\phi),\ \{\zeta_i\} \sim CAR(\lambda_\zeta)$ | 10.4 | 439 |
| 5 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ (\{\phi_i\}, \{\zeta_i\}) \sim MCAR(\rho, \Lambda)$ | 7.9 | 434 |
| 6 | $\mathbf{x}_{ij}^T\boldsymbol{\beta} + \phi_i;\ (\{\phi_i\}, \{\zeta_i\}) \sim MCAR(\rho_\phi, \rho_\zeta, \Lambda)$ | 8.2 | 434 |

Table 14.12 *DIC and $p_D$ values for various competing interval-censored models.*

and primary cancer control. Described more fully by Murray et al. (1998), the data consist of 223 subjects who reside in 53 zip codes in the southeastern corner of Minnesota. The subjects, all of whom were smokers at study entry, were randomized into either a smoking intervention (SI) group, or a usual care (UC) group that received no special antismoking intervention. Each subject's smoking habits were monitored at roughly annual visits for five consecutive years. The subjects we analyze are actually the subset who are known to have quit smoking at least once during these five years, and our event of interest is whether they relapse (resume smoking) or not. Covariate information available for each subject includes sex, years as a smoker, and the average number of cigarettes smoked per day just prior to the quit attempt.

To simplify matters somewhat, we actually fit our spatial cure rate models over the 81 contiguous zip codes shown in Figure 14.10, of which only the 54 dark-shaded regions are those contributing patients to our data set. This enables our models to produce spatial predictions even for the 27 unshaded regions in which no study patients actually resided. All of our MCMC algorithms ran 5 initially overdispersed sampling chains, each for 20,000 iterations. Convergence was assessed using correlation plots, sample trace plots, and Gelman-Rubin (1992) statistics. In every case a burn-in period of 15,000 iterations appeared satisfactory. Retaining the remaining 5,000 samples from each chain yielded a final sample of 25,000 for posterior summarization.

Table 14.12 provides the DIC scores for a variety of random effects cure rate models in the interval-censored case. Models 1 and 2 have only random frailty terms $\phi_i$ with i.i.d. and CAR priors, respectively. Models 3 and 4 add random Weibull shape parameters $\zeta_i = \log \rho_i$, again with i.i.d. and CAR priors, respectively, independent of the priors for the $\phi_i$. Finally, Models 5 and 6 consider the full MCAR structure for the $(\phi_i, \zeta_i)$ pairs, assuming common and distinct spatial smoothing parameters, respectively. The DIC scores do not suggest

| Parameter | Median | (2.5%, 97.5%) |
|---|---|---|
| Intercept | –2.720 | (–4.803, –0.648) |
| Sex (male = 0) | 0.291 | (–0.173, 0.754) |
| Duration as smoker | –0.025 | (–0.059, 0.009) |
| SI/UC (usual care = 0) | –0.355 | (–0.856, 0.146) |
| Cigarettes smoked per day | 0.010 | (–0.010, 0.030) |
| $\theta$ (cure fraction) | 0.694 | (0.602, 0.782) |
| $\rho_\phi$ | 0.912 | (0.869, 0.988) |
| $\rho_\zeta$ | 0.927 | (0.906, 0.982) |
| $\Lambda_{11}$ (spatial variance component, $\phi_i$) | 0.005 | (0.001, 0.029) |
| $\Lambda_{22}$ (spatial variance component, $\zeta_i$) | 0.007 | (0.002, 0.043) |
| $\Lambda_{12}/\sqrt{\Lambda_{11}\Lambda_{22}}$ | 0.323 | (–0.746, 0.905) |

Table 14.13 *Posterior quantiles, full model, interval-censored case.*

that the more complex models are significantly better; apparently the data encourage a high degree of shrinkage in the random effects (note the low $p_D$ scores). In what follows we present results for the "full" model (Model 6) in order to preserve complete generality, but emphasize that any of the models in Table 14.12 could be used with equal confidence.

Table 14.13 presents estimated posterior quantiles (medians, and upper and lower .025 points) for the fixed effects $\boldsymbol{\beta}$, cure fraction $\theta$, and hyperparameters in the interval-censored case. The smoking intervention does appear to produce a decrease in the log relative risk of relapse, as expected. Patient sex is also marginally significant, with women more likely to relapse than men, a result often attributed to the (real or perceived) risk of weight gain following smoking cessation. The number of cigarettes smoked per day does not seem important, but duration as a smoker is significant, and in a possibly counterintuitive direction: shorter-term smokers relapse sooner. This may be due to the fact that people are better able to quit smoking as they age (and are thus confronted more clearly with their own mortality).

The estimated cure fraction in Table 14.13 is roughly .70, indicating that roughly 70% of smokers in this study who attempted to quit have in fact been "cured." The spatial smoothness parameters $\rho_\phi$ and $\rho_\zeta$ are both close to 1, again suggesting we would lose little by simply setting them both equal to 1 (as in the standard CAR model). Finally, the last lines of both tables indicate only a moderate correlation between the two random effects, again consistent with the rather weak case for including them in the model at all.

We compared our results to those obtained from the `R` function `survreg` using a Weibull link, and also to Weibull regression models fit in a Bayesian fashion using the `WinBUGS` package. While neither of these alternatives featured a cure rate (and only the `WinBUGS` analysis included spatial random effects), both produced fixed effect estimates quite consistent with those in Table 14.13.

Turning to graphical summaries, Figure 14.11 maps the posterior medians of the frailty ($\phi_i$) and shape ($\rho_i$) parameters in the full spatial MCAR (Model 6) case. The maps reveal some interesting spatial patterns, though the magnitudes of the differences appear relatively small across zip codes. The south-central region seems to be of some concern, with its high values for both $\phi_i$ (high overall relapse rate) and $\rho_i$ (increasing baseline hazard over time). By contrast, the four zip codes comprising the city of Rochester, MN (home of the Mayo Clinic, and marked with an "R" in each map) suggest slightly better than average cessation behavior. Note that a nonspatial model cannot impute anything other than the "null values" ($\phi_i = 0$ and $\rho_i = 1$) for any zip code contributing no data (all of the unshaded regions in Figure 14.10). Our spatial model however is able to impute nonnull values here, in accordance with the observed values in neighboring regions. ■

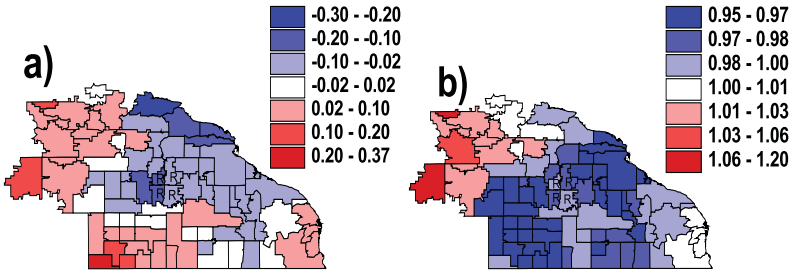


Figure 14.11 *Maps of posterior means for the $\phi_i$ (a) and the $\rho_i$ (b) in the full spatial MCAR model, assuming the data to be interval-censored.*

### 14.6 Exercises

1. The data located at `www.biostat.umn.edu/~brad/data/MAC.dat`, and also shown in Table 14.14, summarize a clinical trial comparing two treatments for *Mycobacterium avium* complex (MAC), a disease common in late-stage HIV-infected persons. Eleven clinical centers ("units") have enrolled a total of 69 patients in the trial, 18 of which have died; see Cohn et al. (1999) and Carlin and Hodges (1999) for full details regarding this trial.

   As in Section 14.1, let $t_{ij}$ be the time to death or censoring and $x_{ij}$ be the treatment indicator for subject $j$ in stratum $i$ ($j = 1, \ldots, n_i$, $i = 1, \ldots, k$). With proportional hazards and a Weibull baseline hazard, stratum $i$'s hazard is then

   $$\begin{aligned} h(t_{ij}; x_{ij}) &= h_0(t_{ij})\omega_i \exp(\beta_0 + \beta_1 x_{ij}) \\ &= \rho_i t_{ij}^{\rho_i - 1} \exp(\beta_0 + \beta_1 x_{ij} + W_i) \, , \end{aligned}$$

   where $\rho_i > 0$, $\boldsymbol{\beta} = (\beta_0, \beta_1)' \in \Re^2$, and $W_i = \log \omega_i$ is a clinic-specific frailty term.

   (a) Assume i.i.d. specifications for these random effects, i.e.,

   $$W_i \overset{iid}{\sim} N(0, 1/\tau) \quad \text{and} \quad \rho_i \overset{iid}{\sim} G(\alpha, \alpha) \, .$$

   Then as in the `mice` example (`WinBUGS` Examples Vol 1),

   $$\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + W_i) \, ,$$

   so that $t_{ij} \sim Weibull(\rho_i, \mu_{ij})$ . Use `WinBUGS` to obtain posterior summaries for the main and random effects in this model. Use vague priors on $\beta_0$ and $\beta_1$, a moderately informative $G(1, 1)$ prior on $\tau$, and set $\alpha = 10$. (You might also recode the drug covariate from (1,2) to (–1,1), in order to ease collinearity between the slope $\beta_1$ and the intercept $\beta_0$.)

   (b) From Table 14.15, we can obtain the lattitude and longitude of each of the 11 sites, hence the distance $d_{ij}$ between each pair. These distances are included in `www.biostat.umn.edu/~brad/data/MAC.dat`; note they have been scaled so that the largest (New York-San Francisco) equals 1. (Note that since sites F and H are virtually coincident (both in Detroit, MI), we have recoded them as a single clinic (#6) and now think of this as a 10-site model.) Refit the model in `WinBUGS` assuming the frailties to have spatial correlation following the isotropic exponential kriging model,

   $$\mathbf{W} \sim N_k(\mathbf{0}, H), \text{ where } H_{ij} = \sigma^2 \exp(-\phi d_{ij}) \, ,$$

   where as usual $\sigma^2 = 1/\tau$, and where we place a $G(3, 0.1)$ (mean 30) prior on $\phi$.

| Unit | Drug | Time | Unit | Drug | Time | Unit | Drug | Time |
|---|---|---|---|---|---|---|---|---|
| A | 1 | 74+ | E | 1 | 214 | H | 1 | 74+ |
| A | 2 | 248 | E | 2 | 228+ | H | 1 | 88+ |
| A | 1 | 272+ | E | 2 | 262 | H | 1 | 148+ |
| A | 2 | 344 | | | | H | 2 | 162 |
| | | | F | 1 | 6 | | | |
| B | 2 | 4+ | F | 2 | 16+ | I | 2 | 8 |
| B | 1 | 156+ | F | 1 | 76 | I | 2 | 16+ |
| | | | F | 2 | 80 | I | 2 | 40 |
| C | 2 | 100+ | F | 2 | 202 | I | 1 | 120+ |
| | | | F | 1 | 258+ | I | 1 | 168+ |
| D | 2 | 20+ | F | 1 | 268+ | I | 2 | 174+ |
| D | 2 | 64 | F | 2 | 368+ | I | 1 | 268+ |
| D | 2 | 88 | F | 1 | 380+ | I | 2 | 276 |
| D | 2 | 148+ | F | 1 | 424+ | I | 1 | 286+ |
| D | 1 | 162+ | F | 2 | 428+ | I | 1 | 366 |
| D | 1 | 184+ | F | 2 | 436+ | I | 2 | 396+ |
| D | 1 | 188+ | | | | I | 2 | 466+ |
| D | 1 | 198+ | G | 2 | 32+ | I | 1 | 468+ |
| D | 1 | 382+ | G | 1 | 64+ | | | |
| D | 1 | 436+ | G | 1 | 102 | J | 1 | 18+ |
| | | | G | 2 | 162+ | J | 1 | 36+ |
| E | 1 | 50+ | G | 2 | 182+ | J | 2 | 160+ |
| E | 2 | 64+ | G | 1 | 364+ | J | 2 | 254 |
| E | 2 | 82 | | | | | | |
| E | 1 | 186+ | H | 2 | 22+ | K | 1 | 28+ |
| E | 1 | 214+ | H | 1 | 22+ | K | 1 | 70+ |
| | | | | | | K | 2 | 106+ |

Table 14.14 *Survival times (in half-days) from the MAC treatment trial, from Carlin and Hodges (1999). Here, "+" indicates a censored observation.*

2. The file `www.biostat.umn.edu/~brad/data/smoking.dat` contains the southeastern Minnesota smoking cessation data discussed in Section 14.5. At each of up to five office visits, the smoking status of persons who had recently quit smoking was assessed. We define relapse to smoking as the endpoint, and denote the failure or censoring time of person $j$ in county $i$ by $t_{ij}$. The data set (already in `WinBUGS` format) also contains the adjacency matrix for the counties in question.
   (a) Assuming that smoking relapses occurred on the day of the office visit when they were detected, build a hierarchical spatial frailty model to analyze these data. Code your model in `WinBUGS`, run it, and summarize your results. Use the DIC tool to compare a few competing prior or likelihood specifications.
   (b) When we observe a subject who has resumed smoking, all we really know is that his failure (relapse) point occurred somewhere between his last office visit and this one. As such, improve your model from part (a) by building an interval-censored version.
3. Consider the extension of the Section 14.4 model in the single endpoint, multiple cause case to the *multiple* endpoint, multiple cause case — say, for analyzing times until diagnosis of each cancer (if any), rather than merely a single time until death. Write down a model, likelihood, and prior specification (including an appropriately specified MCAR distribution) to handle this case.

| Unit | Number | City |
|---|---|---|
| A | 1 | Harlem (New York City), NY |
| B | 2 | New Orleans, LA |
| C | 3 | Washington, DC |
| D | 4 | San Francisco, CA |
| E | 5 | Portland, OR |
| F | 6a | Detroit, MI (Henry Ford Hospital) |
| G | 7 | Atlanta, GA |
| H | 6b | Detroit, MI (Wayne State University) |
| I | 8 | Richmond, VA |
| J | 9 | Camden, NJ |
| K | 10 | Albuquerque, NM |

Table 14.15 *Locations of the clinical sites in the MAC treatment data set.*